

Neugierige Strukturvorschläge

Eine techniknahe interdisziplinäre Einführung in maschinelles Lernen und dessen
Selbstorganisationsprinzipien

Sebastian Harrach



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Inauguraldissertation

Vom Fachbereich Gesellschafts- und Geschichtswissenschaften
der Technischen Universität Darmstadt zur Erlangung eines Grades
des Doktors der Philosophie genehmigte Dissertation

Referenten:

Prof. Dr. Petra Gehring
Prof. Dr. Christoph Hubig

vorgelegt von

Dipl.-Math. Sebastian Harrach
aus Mainz

Eingereicht am: 26.03.2013

Disputation am: 12.06.2013

Darmstadt 2013

D17

Bitte zitieren Sie dieses Dokument als:

URN: [urn:nbn:de:tuda-tuprints-34842](https://nbn-resolving.org/urn:nbn:de:tuda-tuprints-34842)

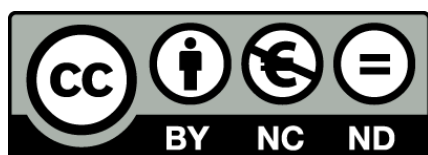
URL: <http://tuprints.ulb.tu-darmstadt.de/3484>

Dieses Dokument wird bereitgestellt von tuprints,


E-Publishing-Service der TU Darmstadt

<http://tuprints.ulb.tu-darmstadt.de>

tuprints@ulb.tu-darmstadt.de



Die Veröffentlichung steht unter folgender Creative Commons Lizenz:
Namensnennung – Keine kommerzielle Nutzung – Keine Bearbeitung 3.0 Deutschland
<http://creativecommons.org/licenses/by-nc-nd/3.0/de/>



Meiner Frau

Ich danke Frauke Harrach für die Entwicklung des Designs und ihren unerschütterlichen Glauben an mich und meine Arbeit.

Ich danke Petra Gehring für das Vertrauen in mein Promotionsvorhaben und die Unterstützung in organisatorisch herausfordernden Kontexten.

Ich danke Christoph Hubig für die große Unterstützung und die Einbeziehung in seine Arbeitsgruppe.

Ich danke Andreas Kaminski für die Einführung in das Institut, die methodische Anleitung und die konstruktiven Gespräche und Rückmeldungen zu verschiedenen Argumentationslinien.

Ich danke Heidi Baetzel und Elke Harrach für das detaillierte Lektorat und den regen Austausch über Interpunktion und Orthographie.

Ich danke Thomas Schuster für die Bekanntmachung mit der Philosophie als Wissenschaft.

Ich danke Fiona vom Rittmeister für die anregenden Spaziergänge.

Inhaltsverzeichnis

Inhaltsverzeichnis	v
1. Einleitung.....	1
1.1. Selbstorganisation und maschinelles Lernen	1
1.1.1. Maschinell lernende Artefakte – MLA	3
1.2. Ein Schritt zurück: Bestimmung des maschinellen Lernens.....	6
1.2.1. Angemessenheit der Rede vom Lernen	7
1.2.2. Technikphilosophische Fragestellungen.....	9
1.3. Konzeptualisierungsstrategie.....	11
1.3.1. Aufbau des ersten Hauptteils	11
1.3.2. Aufbau des zweiten Hauptteils	12
1.3.3. Interdisziplinarität.....	13
2. Erster Hauptteil: Der Blick der Informatik auf maschinelles Lernen	15
2.1. Betrachtungsebenen.....	15
2.2. Unterschiede zwischen lernenden Algorithmen	18
2.2.1. Unterscheidung gemäß erhaltener Rückmeldungen.....	18
2.2.2. Unterscheidung nach Suchstrategien	22
2.2.3. Unterscheidung nach Verwendungszweck	24
2.2.4. Bewertung der Unterscheidungsmöglichkeiten	24
2.3. Klassifizierung nach Lernstrategien.....	26
2.3.1. Überblick der Lernstrategien	26
2.3.2. Lernen von Entscheidungsbäumen.....	28
2.3.3. Evolutionäres Lernen.....	38
2.3.4. Lernen von künstlichen neuronalen Netzen – KNN	51
2.3.5. Instanzenbasiertes Lernen	69
2.3.6. Statistisches Lernen	77
2.3.7. Analytisches Lernen.....	91
2.3.8. Stützvektormethoden	94
2.4. Charakteristik des maschinellen Lernens.....	97
3. Zweiter Hauptteil: Der Blick der Technikphilosophie auf maschinelles Lernen....	98
3.1. Präzisierung der Diskussion am Beispiel von KNN	99
3.1.1. Vorstrukturierte künstliche neuronale Netze.....	99
3.1.2. Prinzipielle Intransparenz von KNN	101
3.1.3. Optionale Zwecklosigkeit der Struktursuche	102

3.2.	Abgrenzung von etablierten Begriffsverwendungen	106
3.2.1.	Überraschung.....	106
3.2.2.	Repräsentation	107
3.2.3.	Experiment.....	111
3.2.4.	Zufallstechnik.....	112
3.2.5.	Unfall	113
3.3.	Ein Schritt zurück: MLA als Informationstechnik und Technik.....	115
3.4.	Suche nach Problembegriffen.....	118
3.4.1.	Quasi-Intentionalität als Ausgangspunkt der Suche	118
3.4.2.	Suche nach Ursprüngen der Irritation.....	125
3.4.3.	Neugier und Vor-Struktur jenseits von Heidegger	135
3.5.	Abgrenzung von aktuellen technikphilosophischen Entwürfen	140
3.5.1.	MLA als naturalisierte Technik	141
3.5.2.	MLA als transklassische Technik.....	152
3.5.3.	MLA als nichttriviale Technik	160
3.5.4.	Fazit der drei Abgrenzungen	166
3.6.	Erzeugung von Welt und maschinelles Lernen	169
3.6.1.	Parallelen des maschinellen Lernens zur Kunst.....	171
3.6.2.	Weisen der Welterzeugung.....	172
3.6.3.	Angemessenheit von Strukturvorschlägen.....	180
3.7.	Zusammenführung der Ergebnisse	182
3.7.1.	Zugang zu Nichtwissen.....	185
3.7.2.	Neugieriges maschinelles Lernen als Technik.....	187
3.7.3.	Welttechnik in der Praxis.....	188
3.7.4.	Ununterscheidbarkeit der entstehenden Weltbezüge	192
4.	Ausblick.....	197
4.1.	Verallgemeinerbarkeit der Suche nach Welttechnik.....	197
4.1.1.	Technoscience als möglicher Suchraum.....	198
4.1.2.	Abduktionsstufen als Mittel zur Suche	199
4.1.3.	Mehrwert einer weiteren Verallgemeinerung.....	201
4.2.	Technikferne Einordnung von MLA.....	203
4.3.	Maschinelles Lernen ohne Computer.....	204
	Abbildungsverzeichnis	207
	Stichwortverzeichnis.....	209
	Literaturverzeichnis	215

1. Einleitung

1.1. Selbstorganisation und maschinelles Lernen

Die Technikphilosophie ist eine Wissenschaft von der Veränderung. Wie sich Gesellschaften oder Teilbereiche durch Technisierung verändern, was der Mensch – kontrafaktisch – ohne Technik wäre und insbesondere wie sich die Technik selbst verändert, zählt zu ihren mit intensiver Neugierde verfolgten Fragen.

(Kaminski 2012, S. 1)

Die Frage, wie sich die Technik selbst verändert, bezieht sich meist auf Durchbrüche in Forschung und Entwicklung, die zur Erschließung neuer Technikfelder führen. Anstelle der Betrachtung einzelner Durchbrüche kann die Frage jedoch auch in Hinblick auf die systematische „Selbstorganisation“ technischer Systeme gestellt werden, das heißt bezogen auf Technik, die sich selbst oder andere Technik adaptiert und reorganisiert. Besonders vielversprechende Veränderungspotenziale können dementsprechend Technik zugeschrieben werden, die sowohl eine Vielzahl von Durchbrüchen in theoretischer und praktischer Hinsicht aufweist, als auch autoadaptive Prinzipien realisiert. Die Untersuchung des Ausmaßes oder überhaupt einer genaueren Bestimmung der auftretenden „Selbstorganisation“ steht in solchen hochdynamischen Technikfeldern allerdings häufig hinter einer Analyse der jüngsten Durchbrüche, der Prognose möglicher zukünftiger Veränderungen und der Abschätzung der Einflüsse des Technikbereiches als Ganzem zurück. Die Frage, wie sich Technik im Rahmen ihres Einsatzes selbst adaptiert und verändert, lässt sich jedoch auch in weniger volatilen und besser verstandenen Bereichen von Technik stellen, etwa in der Informatik. In der Informatik finden sich verschiedene Phänomene, die als Formen von Selbstorganisation beschrieben werden können (Shiffman 2012). Gerade Algorithmen, die auf externe **Sensordaten**¹ zurückgreifen, sind in diesem Zusammenhang besonders interessant. Viele moderne Algorithmen verknüpfen externe Reizkonstellationen, die eine bestimmte Regelmäßigkeit aufweisen, mit speziellen Reaktionen, indem sie, abhängig von bereits erfolgten Eingaben, ihre Struktur verändern. Die entsprechenden autoadaptiven Algorithmen werden innerhalb der Informatik unter der Bezeichnung **maschinelles Lernen** zusammengefasst. Ein besonders interessanter Teilbereich des maschinellen Lernens wird im zweiten Hauptteil der nachfolgenden Darstellungen als die Erstellung **neugieriger Strukturvorschläge** beschrieben und als **Welttechnik** bezeichnet. Die Analyse autoadaptiver Algorithmen und speziell von Welttechnik verspricht dabei nicht nur ein großes Potenzial für die Entdeckung von Veränderungen der Technik, sondern auch die Gelegenheit zu einer präziseren Formulierung dessen, was an den mit „Selbstorganisation“ bezeichneten Phänomenen aus Sicht der Technikphilosophie von zentralem Interesse ist und welche Fragen gestellt werden können oder sollten.

¹ Die Rede von Sensordaten soll sehr allgemein verstanden werden und beschreibt hier und im Folgenden jede Art von Eingabedaten.

Verfahren des machine learning sind die Grundlage von Programmsystemen, die aus „Erfahrung“ lernen, also neues Tatsachen- und Regelwissen gewinnen oder Priorisierungen adaptieren können. Sie sind u.a. auch für die Entdeckung zweckbestimmt relevanter Beziehungen in großen Datenmengen („Data mining“) von großer Bedeutung.

(Görz et al. 2003, S. 13)

Die Technik des maschinellen Lernens bildet in vielen Fällen die Grundlage dafür, dass eine andere Technik als neuartig eingeschätzt wird oder dass die Veränderung einer Technik diagnostiziert wird. Beispiele für maschinelles Lernen sind unter anderem:

- Der Einsatz in der Medizin, wenn Diagnostizierung in Kliniken maschinell unterstützt werden soll (Kononenko 2001). Hier können Krankengeschichten und vergangene Diagnosen von den Systemen bei zukünftigen Diagnosevorschlägen berücksichtigt werden.
- Die Bestimmung von Risikoklassen verschiedenster Arten, sei es bei der Krebsforschung (Kerhet et al. 2009), dem Aktienhandel (Shah 2007) oder der Erstellung von Versicherungstarifen (Dugas et al. 2003). Hier können die genutzten Systeme jeweils aufgrund von Daten der Vergangenheit bestimmte Risikowahrscheinlichkeiten berechnen und diese Berechnungen für jeden neuen Fall aktualisieren. Weiterhin wird in diesen Bereichen versucht, die Strukturmerkmale zu erkennen, die die Risiken am stärksten beeinflusst haben, um diese Merkmale nutzerseitig weiter zu analysieren.
- Die Erstellung von **Recommender-Systemen**. Diese Systeme schlagen den Nutzern Produkte oder Dienste zum Kauf vor. Die Vorschläge basieren auf dem bisherigen Kaufverhalten der jeweiligen Nutzer, aber auch auf dem Kaufverhalten anderer Nutzer. Praktische Beispiele hierzu sind Amazon.com oder der Radiosender Last.fm.
- Die Entdeckung zweckbestimmt relevanter Beziehungen in großen Datenmengen, die im Weiteren wie bei Görz et al. als **Data Mining** bezeichnet wird. In der vorliegenden Arbeit wird der Begriff Data Mining nur verwendet, wenn die zu erlernende Struktur oder die Verwendung der Struktur noch weitgehend unklar ist. Dies dient nur der Vermeidung von Überschneidungen – etwa zu den Recommender-Systemen. Im Rahmen des Data Mining wird maschinelles Lernen eingesetzt, um die am stärksten ausgeprägten Zusammenhänge zwischen Daten zu bestimmen. Diese Zusammenhänge können von einem Menschen auch ohne die Nutzung maschinellen Lernens erkannt und analysiert werden, etwa durch den Einsatz von Algorithmen, die auf dem Konzept der **OLAP-Würfel**² basieren.

Über diese Beispiele hinaus gibt es noch eine sehr große Zahl äußerst unterschiedlicher Anwendungsmöglichkeiten – seien es Pre-Crash-Sensoren für Airbags in Automobilen oder statistische Sprach- und Schrifterkennung. In all diesen Gebieten steht eine Vielzahl von Hindernissen der Lösung der jeweiligen Aufgabe im Weg – seien es ungenaue oder extrem zahlreiche Messwerte, unterschiedliche Messgeräte, veränderliche beobachtete Größen oder ein hoher Zeitdruck. Diese Hindernisse haben gemein, dass sie eine Form von Veränderung

² „Ein OLAP-Würfel [ist ein] gebräuchlicher Begriff zur logischen Darstellung von Daten. Die Daten werden dabei als Elemente eines mehrdimensionalen Würfels [...] angeordnet. Die Dimensionen des Würfels beschreiben die Daten und erlauben auf einfache Weise den Zugriff.“ (Wikipedia Contributors 2012, OLAP-Würfel) Die Pivot-Tabellen in Microsoft Excel stellen ein anschauliches Beispiel für OLAP-Würfel dar.

von der eingesetzten Technik verlangen und dass maschinelles Lernen die Nutzer bei der Überwindung dieser Hindernisse unterstützen kann. Die eingesetzten physischen Objekte, auf denen maschinell lernende Algorithmen implementiert sind, werden im Folgenden **maschinell lernende Artefakte** – kurz **MLA** – genannt. Die neutrale Bezeichnung als **Artefakt** soll betonen, dass das konkrete Objekt sehr unterschiedlich aussehen kann – es ist sogar möglich, maschinelles Lernen ohne einen Computer durchzuführen³. MLA zeichnet aus, dass es ihnen möglich ist, systematisch auf Sensordaten zu reagieren. Die Fähigkeit, auf unbekannte Eingabedaten autoadaptiv zu reagieren, macht die zugrunde liegenden Algorithmen zu lernenden Algorithmen. Einige Algorithmenklassen sind darüber hinaus in der Lage, eine Adaption ihrer eigenen Programmstruktur und sogar eine Adaption der Möglichkeiten zur Adaption ihrer Programmstruktur vorzunehmen. In beiden Fällen bleibt auch ein Algorithmus aus dem Bereich des maschinellen Lernens eine abstrakte Prozessbeschreibung und stellt gerade keine Substanz dar. Entsprechend ist es nicht zielführend, die Möglichkeit eines MLA zur autoadaptiven Reaktion als die Eigenschaft eines speziellen Algorithmus zu denken, da der Algorithmus sonst als Träger dieser Eigenschaft und damit substanzhaft gedacht werden muss. Solch eine Vorgehensweise wäre vergleichbar damit, einen mathematischen Satz als Eigenschaft der mathematischen Logik zu betrachten⁴. MLA werden gezielt als maschinell lernende Artefakte bezeichnet, um die Intuition zu unterstützen, dass die Implementierung in einem physischen Objekt notwendig, dessen konkrete Ausprägung aber nicht relevant ist. Dies soll jedoch, wie bereits erwähnt, nicht dazu führen, dass statt der physischen Ausprägung nun der Algorithmus als die Substanz gedacht wird. Mit der Bezeichnung als Artefakt soll darüber hinaus die Intuition gestärkt werden, dass diese Klasse von Artefakten nicht notwendigerweise auf das maschinelle Lernen oder die Informatik beschränkt ist⁵.

1.1.1. Maschinell lernende Artefakte – MLA

Ein übliches Problem der Technikphilosophie, als Suche nach dem Neuen, liegt darin, dass es im Vorfeld einer Diskussion sehr schwierig ist, zu unterscheiden, ob eine neue Technik das Potenzial hat, eine systematische Neuerung mit sich zu bringen oder ob sich aus einer Diskussion maximal eine neue Perspektive auf bereits bekannte Technik ergeben wird. Im Weiteren soll es darum gehen darzustellen, inwieweit maschinelles Lernen Technik verändert und warum die Veränderungen durch MLA für die Technikphilosophie von besonderem Interesse sind. Angesichts der weiten Verbreitung maschinellen Lernens und des großen Funktionsumfangs dieser Technik stellt sich zunächst die Frage, was einer genauen Analyse der konkreten Artefakte bisher im Wege stand. Die beiden größten Schwierigkeiten, die eine solche Diskussion unattraktiv erscheinen lassen, sind die Überschätzung der Selbstständigkeit maschinell lernender Artefakte und deren fehlende Wahrnehmbarkeit. Beide Schwierigkeiten werden im Folgenden kurz skizziert.

Selbstständigkeit von MLA

MLA werden durch **Autoadaptionsprozesse** konstituiert. Eine aus einem Missverständnis dieser Autoadaptionsprozesse resultierende Fehleinschätzung liegt darin anzunehmen, dass es

³ Details finden sich im abschließenden Abschnitt 4.4.3.

⁴ Beide Betrachtungsweisen sind gleichermaßen sowohl formal zulässig als auch nicht zielführend.

⁵ Eine Bezeichnung der betrachteten Objekte als maschinell lernende **Implementierungen** hätte den gegenteiligen Effekt.

sich um Artefakte handelt, die selbsttätig Muster in Daten erkennen und daraus, unabhängig von menschlichen Einflüssen, Konzepte oder Modelle erstellen beziehungsweise diese Modelle eigenständig in den Rohdaten entdecken. Tatsächlich entstehen die von einem MLA präsentierten Konzepte entweder schon bei der Konzeption des **Lösungsraums**, das heißt der Festlegung, welche Elemente die Lösung beziehungsweise das Ergebnis auszeichnen sollen beziehungsweise dürfen, oder erst nachträglich im Rahmen einer Interpretation der Ergebnisse durch den Menschen. Im Fall einer nachträglichen Interpretation können im Rahmen des Autoadaptionsvorgangs durchaus Strukturen entstanden sein, die neben einem systematischen auch einen unabhängig vom Nutzer entstandenen Anteil enthalten. Diese Anteile der entstandenen Struktur werden jedoch erst im Nachhinein als Konzepte interpretiert beziehungsweise wahrgenommen.

Problematisch wird diese Überschätzung von MLA, weil die Artefakte dadurch in Diskussionen auftauchen, in denen sie keinen Platz haben. Entsprechend stehen die Spezifika und technikphilosophisch interessanten Aspekte von MLA in solchen Projekten nicht im Fokus. Darüber hinaus sinkt der empfundene Bedarf, maschinelles Lernen technikphilosophisch aufzuarbeiten, da es häufig als Randphänomen von Technikbereichen wahrgenommen wird, die noch sehr viel Unerforschtes enthalten. Die Entwicklung des Begriffes der Welttechnik im zweiten Hauptteil wird einen Erklärungsansatz liefern, warum gerade noch unbekannte Technikbereiche in besonderem Maße mit maschinell lernenden Artefakten erforscht werden.

MLA in der Wahrnehmung der Nutzer

Die zweite Schwierigkeit liegt darin, dass Menschen, die MLA einsetzen, meist kein Verständnis dafür haben, welchen Typ von Artefakt sie nutzen, beziehungsweise was es ist, das dieses Artefakt ihnen als („Lern-“)Ergebnis anbietet oder gar was an dem Ergebnis unabhängig vom Menschen erzeugt wurde. Die fehlende Wahrnehmbarkeit maschinell lernender Artefakte bezieht sich nicht in erster Linie auf deren beobachtbare Funktion, sondern auf die Hypothesen über das MLA als Artefakt. Schon vor dem Versuch, eine Funktionsweise zu beobachten, müssen Nutzer Hypothesen aufstellen, was für eine Art von Artefakt vorliegt – so erfordern etwa technische Werkzeuge, Kunstwerke oder Lebewesen jeweils eine spezifische Weise des Nachdenkens über so etwas wie eine Funktionsweise oder Einsetzbarkeit. Wenn der Nutzer ein MLA als Artefakt überhaupt nicht wahrnimmt, ist die Situation verwandt, allerdings verschiebt sich in diesem Fall das Problem in Richtung der Frage nach der Steuerung beziehungsweise Regelung des Artefaktes durch den jeweiligen Nutzer oder den Entwickler des MLA. Diesbezüglich weisen MLA mit ihrer Autoadaptivität eine Besonderheit gegenüber anderen Formen nicht wahrgenommener oder prinzipiell nicht wahrnehmbarer Technik auf, wie sie etwa im **Ubiquitous Computing** eingesetzt werden. Die Autoadaptivität von MLA kann sehr unterschiedliche Ausprägungen aufweisen, verschiedene Aspekte des Artefaktes betreffen und sich in ihrer Ausprägung verändern. Gleichzeitig basieren vielen Prozesse, die von ihren Nutzern dem maschinellen Lernen zugeordnet werden, nicht oder nur zu sehr nachgeordneten Teilen auf dieser Technik. Ein Beispiel für solch einen voreilig zugeordneten Prozess ist die Erstellung und Wartung von Spamfiltern. Hier versucht ein menschliches Gegenüber aktiv, die Technik zu umgehen und zu behindern, entsprechend müssen in diesem Fall wesentliche Prozessschritte von Sicherheitsexperten übernommen werden (Weimer 2010).

Insgesamt werden zum einen im Rahmen der Nutzung der MLA Erwartungshaltungen und Technikbegriffe zugrunde gelegt, die in Hinsicht auf andere Artefakte, Maschinen und

Systeme geprägt wurden. Zum anderen arbeiten MLA selbsttätig und verwenden Methoden, die zum Teil notwendigerweise außerhalb des Nutzer-Verständnisses liegen. Entsprechend beschäftigen sich technikphilosophische Betrachtungen von Computern oder Computerprogrammen in der Regel nicht mit maschinellern Lernen als tatsächlich existierender Technik und Grundlage bestehender Artefakte. Meist wird eine Betrachtung des Charakters der eingesetzten Algorithmen wie etwa maschinellen Lernstrategien als wenig relevant vernachlässigt – dies für den jeweiligen Fokus durchaus zu Recht. Allerdings begegnet die entsprechende Betrachtungsweise dem maschinellen Lernen so nie explizit, da dessen Beitrag bei den meisten Artefakten nicht sofort offensichtlich ist und vielfach Autoadaptionsvorgänge schon lange abgeschlossen sind, wenn die Artefakte oder die Auswirkungen von Artefakten betrachtet werden.

Kurzübersicht des Forschungsstandes

Trotz der großen praktischen Bedeutung dieser Form der Technik ist der aktuelle Forschungsstand der Technikphilosophie zum maschinellen Lernen unbefriedigend.

Die Umsetzung wichtiger Leitmotive des Ubiquitous Computing wie Kontextsensitivität und -adaptivität, Autonomie, Selbstorganisation und Smartness basieren auf [maschinellern Lernen]. Maschinelle Lernverfahren können zwar als Fortsetzung des schwachen Artificial Intelligence Paradigmas verstanden werden, gleichwohl erhalten sie keineswegs eine vergleichbare Aufmerksamkeit in der technikphilosophischen Literatur.

(Kaminski 2012, S.2)

Wie Kaminski ausführt, wurden in thematischer Nähe zu maschinellern Lernen technikphilosophische Bestimmungsversuche bezüglich des IT-Paradigmas (Hubig 2008; Wiegelerling et al. 2008; Kaminski et Winter 2011; Wiegelerling 2011) und der daraus resultierenden Veränderungen unternommen. Weiterhin wurde sowohl durch die Philosophie (Eco 1983; Kaminski et Harrach 2010) als auch durch die Informatik (Kakas et al. 1993; Flach 1994; Kakas et al. 1998; Mooney 2000) die Frage nach dem Ausmaß und der Art der abduktiven Fähigkeiten maschinell lernender Artefakte gestellt. Eine umfassende Diskussion der Potenziale maschinellen Lernens wurde bisher jedoch allenfalls von der Informatik selbst angestoßen (Mattern 2003; Fleisch et Mattern 2005; Mattern 2007).

Aktuelle Betrachtungen von Computern, sei es als System (Hippe 2012) oder als Ding (Alpsancar 2012), begegnen dem maschinellen Lernen meist nicht oder nur als Randphänomen, da vielfach Autoadaptionsprozesse zum Beobachtungszeitpunkt schon vorüber sind oder ablaufende Vorgänge nicht ohne Weiteres als maschinellern Lernen identifiziert werden können.

1.2. Ein Schritt zurück: Bestimmung des maschinellen Lernens

Erst wenn technikphilosophisch besser verstanden oder beschreibbar ist, was maschinelles Lernen konstituiert und ermöglicht, wird eine Analyse, Diskussion und Verortung des maschinellen Lernens möglich. Eine zentrale Stärke von MLA, die durch ein besseres technikphilosophisches Verständnis sichtbar gemacht und weiter analysiert werden soll, liegt in deren Fähigkeit, **Rohdaten** unvoreingenommen zu registrieren. Im Gegensatz zum Menschen werden *spezielle* MLA weder durch implizites Vorwissen noch durch implizite Vorannahmen über die Herkunft der Rohdaten⁶ bei deren Wahrnehmung beeinflusst. Diese Stärke kann vom Nutzer bewusst eingesetzt werden und sie kann nur im Zusammenhang mit Artefakten auftreten, denen sich eine gewisse Form von „Selbstorganisation“ zuschreiben lässt. Der Versuch einer systematischen Eingrenzung des maschinellen Lernens beziehungsweise einer Klärung der Frage, was maschinelles Lernen konstituiert, erfolgt im Folgenden auf Basis der Begriffe des Strukturvorschlages und des Lernens. Die Frage nach der Angemessenheit der Rede von Lernvorgängen wird hierbei besonders im Fokus stehen.

Der Bestimmung dessen, was im Weiteren unter maschinellem Lernen verstanden wird, soll mit einer kurzen Klarstellung begonnen werden, was *nicht* darunter verstanden wird. Die zentrale Aussage dieser Vorüberlegung besteht darin, dass allein die *Ergebnisse* maschinellen Lernens noch kein maschinelles Lernen sind. Wie sehr MLA mit den Ergebnissen ihrer Autoadaptionsprozesse identifiziert werden, wird unter anderem daran deutlich, dass MLA und mitunter auch die ihnen zugrunde liegenden Strategien nach der Struktur ihrer Ergebnisse benannt werden, etwa „Entscheidungsbaum“ oder „künstliches neuronales Netz“. Ein bereits angedeutetes Beispiel für ein Resultat maschinellen Lernens besteht in der automatischen Generierung von Empfehlungen, wie den Produktempfehlungen von Amazon oder den Musikempfehlungen des Online-Radiosenders Last.fm. Diese Empfehlungen als solche stellen kein maschinelles Lernen dar, sondern sind nur das Resultat eines Autoadaptionsprozesses. Entsprechend ist eine Manipulation dieser Empfehlungen – etwa eine manuelle Priorisierung oder eine Interpretation, welche Konzepte die Empfehlung erklären könnten – nicht direkt Teil des maschinellen Lernens. Nutzerreaktionen, die nach dem Abschluss eines Lernvorgangs auftreten, erfolgen nur in Ausnahmefällen im Hinblick auf den eigentlichen Autoadaptionsprozess. Im Normalfall setzt sich der Nutzer ausschließlich mit dem Resultat eines solchen Prozesses auseinander. Das angebotene Resultat wird dabei entweder verworfen oder akzeptiert und operationalisiert. Das MLA besteht im Beispiel der Produktempfehlungen aus der IT-Architektur und der zugehörigen Hardware im Hintergrund, die die Einkaufsdaten speichert, ausliest und die nutzerübergreifenden sowie die nutzerspezifischen Empfehlungsschemata aktualisiert. Die aus den Empfehlungsschemata resultierenden Empfehlungen sind **Strukturvorschläge** des MLA an die Nutzer, die von diesen eingesetzt oder abgelehnt werden können. Der lernende Teil eines MLA besteht entsprechend in einem sehr unterschiedlich implementierbaren Gebilde⁷, das Strukturvorschläge erstellt. Die Feststellung, dass sehr unterschiedliche Implementierung zu Lernergebnissen führen, die dieselben Strukturvorschläge erzeugen, zeigt, dass die vom Nutzer wahrgenommenen Strukturvorschläge keinen Bestandteil der zugrunde liegenden MLA darstellen. Knapp

⁶ Hier ist nicht die Annahme des Bestehens einer Quelle der Rohdaten sondern Annahmen über die konkrete Ausprägung der Quelle und der Art des registrierten Reizes gemeint.

⁷ Die unterschiedlichen Möglichkeiten zur Implementierung dieses Gebildes werden im ersten Hauptteil aufgezeigt. Die Rolle des Gebildes innerhalb eines MLA wird im zweiten Hauptteil als die einer Transformationsfunktion identifiziert und analysiert.

formuliert können Empfehlungsalgorithmen und die resultierenden, Strukturvorschläge erzeugenden **Recommender-Systeme** zwar Resultate von Autoadaptionsprozessen sein, aber auch wenn sie einen solchen Ursprung aufweisen, stellen sie nicht das MLA als solches dar. Nutzer setzen sich entsprechend nur selten in ein Verhältnis zum MLA, sondern meist nur zu den von Lernergebnissen erstellten Strukturvorschlägen. Dies unterscheidet MLA nicht von anderen Typen von Technik, da sehr häufig eine **Nutzeroberfläche** zur Verfügung steht, die die eigentliche Funktionsweise mehr oder weniger stark verbirgt. Die Diskussion von MLA leidet jedoch besonders unter der Verwechslung von Funktion und Nutzeroberfläche, wie im zweiten Hauptteil ausführlich dargestellt wird.

Im Bisherigen und Weiteren war und wird stets von **Nutzern** die Rede sein, unabhängig davon wie der betreffende Mensch sich im Einzelfall konkret zu den MLA beziehungsweise den Rohdaten verhält⁸. Ob ein Nutzer die Daten erhoben hat, ob er die Daten nur besitzt, sie betrachtet, interpretiert oder zu einem Zweck nutzen möchte, wird im jeweiligen Kontext explizit diskutiert, wenn es relevant ist. Auch die Frage, ob ein Nutzer sich des Einsatzes des MLA bewusst ist, spielt bei der Bezeichnung keine Rolle und die Antwort ergibt sich aus dem jeweiligen Kontext der Analyse, wenn sie eine Rolle spielt. Der Nutzer wird durch den Einsatz des MLA zum Nutzer. Der Ausdruck **maschinell lernende Artefakte** bezeichnet Artefakte beliebiger Realisierung und physischer Ausprägung, die mitgedacht werden muss und die die praktische Umsetzung maschinellen Lernens ermöglicht. Maschinelles Lernen wiederum beschreibt Algorithmen, die Rohdaten erhalten und darin Reizkonstellationen identifizieren, die eine bestimmte Regelmäßigkeit aufweisen. Diese Reize wiederum werden mit speziellen Reaktionen verknüpft, indem die Algorithmen abhängig von den erhaltenen Eingabedaten ihre Struktur verändern. Maschinelles Lernen bezieht sich im Folgenden konkret auf die Gesamtheit der Gruppe unterschiedlich lernender Algorithmen, die im ersten Hauptteil konkret beschrieben werden. Maschinelles Lernen beschreibt also Algorithmen und damit ein Teilgebiet der Informatik. Maschinell lernende Artefakte entstammen jedoch nicht per se der Informatik, so wie das Motorsteuergerät eines Automobils nicht Teil der Informatik ist. MLA sollten dementsprechend nicht auf ihren algorithmischen Hintergrund reduziert werden, wobei die Diskussion der Unterscheidung zwischen Software und Hardware oder Sinnhaftigkeit der Rede von Software überhaupt an dieser Stelle ausgeklammert werden kann. All dies paraphrasiert die bisherige Verwendung des Begriffes vom maschinellen Lernen. Es verbleibt die Frage, inwiefern hier von einem Lernvorgang gesprochen werden kann.

1.2.1. Angemessenheit der Rede vom Lernen

Wenn von maschinellern Lernen die Rede ist, ist nicht Lernen mit Maschinen oder E-Learning gemeint, sondern die Musterbildung autoadaptiver Systeme. Das Augenmerk liegt auf Algorithmen, die ihre Struktur auf Basis von Sensordaten auf eine bestimmte Weise adaptieren. Die so beschriebenen MLA realisieren eine maschinelle Ausprägung von assoziativem Lernen (Drepper 2010, S. 77) beziehungsweise von Reiz-Reaktions-Lernen (Raimer 2002, S. 66), das heißt, begrifflich ist es durchaus vertretbar, solche Algorithmen als lernend zu beschreiben. Darüber hinaus scheint es auch inhaltlich plausibel zu sein, von Lernvorgängen zu sprechen, wenn Algorithmen betrachtet werden, die etwa ein Computerspiel ohne menschliche Anleitung rein auf Basis der Lektüre des Handbuchs oder

⁸ Weiter wurde und wird bezüglich der Nutzer zur einfacheren Lesbarkeit lediglich die männliche Anredeform gewählt.

der Betrachtung eines Erklärungsvideos „erlernen“ und erfolgreich gegen menschliche Spieler antreten können (Branavan et al. 2011; Kaiser 2012). Gleichwohl ist diese Beschreibung nicht unproblematisch, denn eine Aufnahme und Speicherung von Sensordaten nimmt heute so gut wie jeder komplexere Algorithmus vor – etwa wenn das zugehörige Artefakt in Betrieb genommen wird und vor der ersten Nutzung erst konfiguriert werden muss. Entsprechend muss unausgesprochen der Fokus der Arbeitsweise des Artefaktes auf der Verarbeitung von Sensordaten liegen, damit das Artefakt als ein MLA eingestuft wird. Darüber hinaus kann ein Artefakt seinen Status einbüßen, wenn es die Verarbeitung und Speicherung von Sensordaten einstellt, insofern wäre ein Artefakt, dass nur im Rahmen der Ersteinrichtung nach dem Namen des Nutzers fragt, zwar formal ein MLA, allerdings nur in den ersten Minuten seines Einsatzes. Diese Problematik soll hier zwar aufgezeigt werden, sie wird im Folgenden jedoch ausgeklammert. Die Frage, ob die Rede von einem Lernvorgang angemessen ist oder ob gar maschinelles Lernen einem menschlichen Lernen entspricht, wird im Weiteren nicht aufgegriffen und allenfalls implizit beantwortet, indem die Verwendung des Begriffs des Lernens vermieden und damit als nicht notwendig betrachtet wird. Die Diskussion wird von **Autoadaptionsprozessen** und nicht von Lernprozessen sprechen und die beiden Begriffe werden insofern als synonym betrachtet, als dass die Autoadaptation im Weiteren immer basierend auf Sensordaten gedacht wird⁹.

Die Anpassung an Sensordaten im Rahmen einer Autoadaptation kann zu einer Vielzahl von Ergebnissen führen und es stellt sich die Frage was es bedeuten kann, dass ein solches Ergebnis auch relevant ist. Im weiteren Verlauf dieser Diskussion wird auf die Rolle der Ergebnisse von Autoadaptionsprozessen noch intensiv eingegangen. Vorweg kann festgehalten werden, dass innerhalb der Informatik häufig davon gesprochen wird, dass ein maschinell lernendes Artefakt durch den Autoadaptionsvorgang eine **Funktion** erlernt. Das Erlernen einer Funktion beschreibt in diesem Zusammenhang, dass das betrachtete MLA nach Abschluss des Autoadaptionsprozesses Reaktionsmuster aufweist, die eine im Vorfeld festgelegte Funktionalität realisieren. Die Anforderungen an die Präzision mit der die Funktion erfüllt wird, hängen stark vom jeweiligen Kontext ab und können sich je nach Situation auch verändern. Weitere Tests können beispielsweise ergeben, dass das MLA nur in einer speziellen Umgebung erfolgreich ist oder dass das MLA die Funktion wieder verlernt hat, nachdem weitere Eingabedaten erfolgt sind. Ein MLA hat im Normalfall kein Konzept davon, ob und wie es eine erlernte Funktion für die Zukunft konservieren kann. Wenn ein Nutzer ein MLA eine Funktion erlernen lassen will, gilt das Gleiche. Hierbei ist es hilfreich sich vor Augen zu führen, dass auch ein völlig zufälliger Autoadaptionsprozess aus Sicht des Nutzers der Erlernung einer Funktion entspricht, wenn das MLA zu einem bestimmten Zeitpunkt die Funktion erfüllt und der zufällige Prozess daraufhin abgebrochen beziehungsweise eingefroren wird. Ob eine vergleichbare Situation bei Menschen auch auftritt und wie sich die Kontexte unterscheiden, steht hier nicht im Fokus. Es tritt jedoch häufig ein interessanter Bruch in der Beschreibungslogik auf, der kurz aufgeklärt werden soll. Angenommen, ein Nutzer möchte lernen Schach zu spielen und setzt ein MLA ein. In diesem Fall würde der Nutzer davon sprechen, dass das MLA an seiner Stelle erlernt das Spiel zu spielen und es scheint der Versuch einer direkten Übertragung eines menschlichen Lernvorgangs auf ein Artefakt stattgefunden zu haben. Allerdings liegt für die meisten Fälle lediglich eine unscharfe Formulierung vor, der Nutzer möchte typischerweise nicht nur

⁹ Eine sowohl aus Sicht der Technikphilosophie als auch aus Sicht der Informatik gut lesbare Einführung in die Begrifflichkeiten des Lernens mit direktem Bezug zu maschinellern Lernen bietet Tönnsen (Tönnsen 2007, S. 7ff).

erlernen, welche Figuren an welche Stellen bewegt werden sollten, um maximale Siegchancen zu haben. Das implizite Ziel ist es, auf einer abstrakten Ebene zu verstehen, warum die Figur an eine bestimmte Stelle gezogen werden muss und die reine Fähigkeit, Figuren erfolgreich zu bewegen, wird selten mit der Fähigkeit Schach zu spielen gleichgesetzt. Wenn diese Gleichsetzung erfolgt, kann ein MLA auch erlernen Schach zu spielen, allerdings hat der Nutzer in diesem Fall keinen Nutzen von dem Lernergebnis, da er die Fähigkeit weiterhin nicht besitzt. Aus demselben Grund würde kein Nutzer ein MLA Fahrrad fahren erlernen lassen, wenn er selbst gern Fahrrad fahren könnte. Zusammengefasst spielt bezüglich des menschlichen Lernens in der hier angestrebten Diskussion lediglich die Frage eine Rolle, inwieweit sich der Status von MLA als Technik verändert, wenn Nutzer ihnen Lernvorgänge menschlicher Ausprägung zuschreiben¹⁰. Diese Attribuierung beeinflusst die Bedeutung, die maschinellern Lernen beigemessen wird und die genannte Fehlwahrnehmung wesentlicher Eigenschaften von MLA.

1.2.2. Technikphilosophische Fragestellungen

Die Vielschichtigkeit der Frage, inwiefern es sich bei MLA um Technik handelt, wird im zweiten Hauptteil ausführlich diskutiert. Ein erster Eindruck ergibt sich am besten durch das Anreißen einer möglichen technikphilosophischen Perspektive. Die Perspektive von Kaminski (Kaminski 2012) eignet sich besonders gut für eine kurze Vorbetrachtung, da sie sich explizit mit maschinellern Lernen beschäftigt.

In allen betrachteten Fällen besteht der Lernvorgang aber darin, für gegebene Daten eine Funktion zu finden. Diese Funktion entspricht [...] einem Modell, das die Ordnung der gegebenen Daten beschreibt. Diese Ordnung weist einen Zeitindex auf, es handelt sich um eine Art Hypothese, dass das Modell nicht nur die vergangenen, sondern auch die zukünftigen Daten beschreibt. Aufgrund dieser Zeitlichkeit ist die Modellbildung dynamisch, was die Dynamik der Transformation kennzeichnet. Einige Unterschiede zwischen den genannten Lernstrategien gehen darauf zurück, wie dies erfolgt: Der Hypothesenraum wird von den verschiedenen Lernstrategien anfänglich auf unterschiedliche Weise entworfen bzw. später dann unterschiedlich angepasst.

(Kaminski 2012, S. 12)

Auch wenn man der Beschreibung Kaminskis vollständig zustimmt, ergibt sich eine Vielzahl von Anschlussfragen:

- Wird die Funktion in den Rohdaten gesucht oder für die Rohdaten konstruiert?
- Was kommt als Modell in Frage und ist das Modell intelligibel? Wie ist mit konfligierenden Modellen umzugehen?
- Sind die relevanten Daten vollständig gegeben und was unterscheidet sie von weiteren vorliegenden Rohdaten¹¹?
- Wie wird überprüft, ob das Modell eine weitreichende Aussagekraft aufweist?
- Wie unterscheiden sich diese Lernvorgänge von Optimierungsalgorithmen?

¹⁰ Dieser Punkt wird insbesondere in Abschnitt 3.4 diskutiert.

¹¹ Mit anderen Worten: sind die Daten repräsentativ?

Diese Fragen sind nicht alle von gleicher Brisanz und Relevanz, sollen jedoch jeweils andeuten, dass zumindest prinzipiell sehr unterschiedliche Klassen von maschinellern Lernen denkbar sind und dass diese Klassen aus technikphilosophischer Sicht in Bezug auf unterschiedliche Fragestellungen interessant sind. Der grundlegende Standpunkt, der von Kaminski in der zitierten Passage angedeutet und im restlichen Text und in Vorträgen ausgearbeitet wurde, lautet, dass maschinelles Lernen als Selbstorganisation von Repräsentationssystemen verstanden werden kann. Wie sich im Weiteren zeigen wird, ist das für viele, aber nicht für alle Klassen maschinellen Lernens eine valide Aussage. Kaminski selbst spricht von den Unterschieden zwischen Lernstrategien und genau diese Unterschiede sind für eine systematische Analyse des maschinellen Lernens und dessen Bedeutung für die Technikphilosophie nicht zu vernachlässigen. Aufbauend auf einer solchen Unterscheidung von Klassen maschinellen Lernens lässt sich etwa analysieren, welche Klassen maschinellen Lernens sich als Repräsentationssysteme beschreiben lassen und welche einen anderen Ansatz erfordern.

1.3. Konzeptualisierungsstrategie

Die Hauptthese dieser Einleitung lautet, dass eine Aufarbeitung, Diskussion und Verortung des maschinellen Lernens erst möglich ist, wenn technikphilosophisch präziser verstanden und beschreibbar wird, was maschinelles Lernen konstituiert und ermöglicht. Die Rede von Präzision bezieht sich dabei durchaus auf eine mathematische Präzision, also den Einsatz und die Aufbereitung der präzisen Fachbegriffe der Informatik, zumindest soweit dies notwendig ist. Die Konzeptualisierungsstrategie zielt darauf ab, maschinelles Lernen begrifflich zu erfassen. Die resultierende Strategie muss die konstruktive Konfrontation der Phänomene des maschinellen Lernens mit einer anderen Sprache sicherstellen. Zu diesem Zweck werden nach Klärung der Grundlagen verschiedene technikphilosophische Perspektiven auf ihren Mehrwert für die Diskussion überprüft, es wird eine Annäherung in Form einer Ökonomie der Gewinne und Verluste genutzt, die aufgrund der Vorarbeit präzise genug sein wird, keine geisteswissenschaftliche Außenansicht darzustellen. Perspektiven werden getestet und die Verständniszuwächse werden beim Test der jeweils nächsten Perspektive genutzt. Der Versuchsaufbau der Analyse besteht weiter darin, Begriffe aus dem im Zusammenhang mit maschinellem Lernen – zunächst – fremd erscheinenden Kontext des Weltbegriffes zu verwenden, um damit das Problem einzukreisen. Eine virulente Frage bei dieser Art der Analyse wird dementsprechend darin bestehen, immer wieder zu bestimmen, inwiefern Begriffe, die ursprünglich aus einem anderen Kontext stammen, bestimmte Typen von technischen Artefakten modellieren oder rekonstruieren können.

Maschinelles Lernen liegt in einem Spannungsverhältnis zwischen philosophischen und technischen Technikbegriffen. Dieses Spannungsverhältnis eröffnet in Kombination mit der genannten Konzeptualisierungsstrategie einen Suchraum für die Identifikation eines Technikkonzepts, das MLA als neuartige Form von Technik besser darstellbar macht. Die Diskussion dieses Suchraumes führt im Schlussteil der vorliegenden Analyse auf einen Teilbereich des maschinellen Lernens, der als systematisch neuer Beitrag zur Vielfalt menschlicher Technik identifiziert und am Ende des zweiten Hauptteils als **Welttechnik** bezeichnet wird. Die dem eingesetzten Technikkonzept zugrunde liegende Haltung zur Welt ist auf die Erzeugung von Optionen von Welt ausgerichtet und besagt, dass der Mensch beim Versuch der Erschließung von sehr komplexen oder sehr zahlreichen Rohdaten Herausforderungen gegenüber steht, die technisch gelöst werden können.

1.3.1. Aufbau des ersten Hauptteils

Im ersten Hauptteil wird das maschinelle Lernen als Teil der Informatik betrachtet und mathematisch präzise aufgearbeitet. Das Ziel ist, ein technikphilosophisches Verständnis der technischen Hintergründe zu ermöglichen und so die Basis für einen interdisziplinären Austausch zu schaffen. Zu diesem Zweck sollen die technischen Grundlagen des maschinellen Lernens innerhalb der Informatik annähernd verlustfrei von der Sprache der Mathematik in die Sprache der Philosophie übersetzt werden.

Einen Algorithmus als maschinelles Lernen zu klassifizieren ist noch recht unkonkret, etwa so wie einen Transrapid als Transportmittel zu beschreiben. Innerhalb des maschinellen Lernens sind sehr unterschiedliche Ansätze zu finden, wie ein Algorithmus beziehungsweise ein MLA Strukturen in Eingabereizen finden und wie der Autoadaptionsprozess ablaufen kann. Diese unterschiedlichen Ansätze lassen sich zu unterschiedlichen **Lernstrategien** zusammenfassen. Die Unterscheidung und Diskussion dieser Lernstrategien ist der zentrale Inhalt des ersten

Hauptteils. Zur Vorbereitung dieser Diskussion wird zuerst dargestellt, in welchen Aspekten sich MLA prinzipiell unterscheiden können. Eine wichtige Dimension, nach der MLA unterschieden oder bewertet werden können, ist etwa ob und auf welche Weise durch ein MLA ein Lernerfolg festgestellt werden kann. Die neutralere Formulierung des Autoadaptionsprozesses führt intuitiv auf das zugrunde liegende Problem, dass eine Adaption prinzipiell beliebig ablaufen kann und sich nicht unbedingt Strukturen herausbilden müssen. Eine mögliche Lösung besteht darin, dass ein MLA eine Form von Rückmeldung darüber erhalten kann, ob seine Ausgabe vor oder nach dem letzten Autoadaptationsschritt zu bevorzugen ist. Solch eine Rückmeldung stellt sicher, dass der Autoadaptation eine Systematik zugrunde liegt. Wenn keinerlei Systematik vorliegt, entsteht ein Algorithmus, der annähernd zufällige Ergebnisse produziert. Eine weitere vergleichsweise allgemeine Unterscheidungsdimension zur Unterscheidung von MLA liegt in der Betrachtung der Charakteristik der Zwischenhalte des Autoadaptionsprozesses. Eine Autoadaptation kann kleinschrittig vorgehen und Strukturen hervorbringen, die bisherigen Adaptionsergebnissen ähneln, oder der Prozess kann ausgehend von dem letzten Ergebnis sehr stark auf bereits erlernte Strukturen zurückgreifen und sich im nächsten Schritt weit von dem Ergebnis entfernen. Die beiden Unterscheidungsdimensionen wurden herausgegriffen, da sich schon bei einer kurzen Betrachtung der Eindruck aufdrängt, dass die Dimensionen nicht unabhängig voneinander sind. Das ist eine wesentliche Beobachtung und betont noch einmal, dass es wichtig ist, die technischen Hintergründe tatsächlich zu einem gewissen Grad zu durchdringen, bevor eine Diskussion der Verortung des maschinellen Lernens als neue oder bereits bekannte Technik erfolgen kann.

Diese prinzipiellen Unterscheidungen zwischen lernenden Algorithmen sind allerdings nicht die Grundlage der Kriterien, nach denen die Algorithmen zu Klassen von Lernstrategien zusammengefasst werden. Diese Kriterien sind wesentlich weniger technisch, sie finden sich zu großen Teilen in den den Algorithmen zugrunde liegenden Intuitionen. Entsprechend sind technische Mischformen der im ersten Hauptteil erstellten Klassen von Lernstrategien in der Praxis vollkommen üblich. Das ist nicht sofort verständlich und sollte zum jetzigen Zeitpunkt lediglich im Hinterkopf behalten werden.

1.3.2. Aufbau des zweiten Hauptteils

Die Hauptmotivation des zweiten Hauptteils und der Analyse insgesamt besteht darin, dass technikphilosophisch zunächst weder sinnvoll formulierbar ist, was MLA ausmacht, noch inwiefern ein Mensch überhaupt mit einem MLA in Relation stehen kann. Die Nebenmotivation des zweiten Hauptteils besteht darin, aus der Detailanalyse des maschinellen Lernens Fragen zu generieren und diese als Anknüpfungspunkte für auf der vorliegenden Arbeit aufbauende Analysen auszuzeichnen. Die zur Erreichung dieser beiden Ziele notwendige Diskussion wird durch die Grundlagenarbeit des ersten Hauptteils bereits vorbereitet und begonnen. Im zweiten Hauptteil werden auf die Grundlagenarbeit aufbauend die wesentlichen Charakteristika des maschinellen Lernens aus technikphilosophischer Perspektive analysiert und als Technik verortet. Zwar wird die Aufschlüsselung des maschinellen Lernens als einem Gebiet in Teilgebiete mit unterschiedlichen Eigenschaften im ersten Hauptteil bereits aus der Perspektive der Informatik dargestellt, dennoch baut die technikphilosophische Unterteilung in ihrer Systematik nicht unmittelbar auf dieser Unterteilung auf. Der zweite Hauptteil beginnt entsprechend wieder mit einer Gesamtbetrachtung des maschinellen Lernens, allerdings in diesem Schritt bereits mit dem

Hintergrundwissen bezüglich der einzelnen Lernstrategien und dem Detailverständnis möglicher technischer Unterscheidungen.

Die Methodik im zweiten Hauptteil wird, wie schon angedeutet, in einem Ablaufen von Sackgassen bestehen, bei dem jedoch nur diejenigen Sackgassen vorgestellt und diskutiert werden, aus denen etwas gewonnen werden kann – seien es Negativbefunde, die den Suchraum verkleinern, in dem maschinelles Lernen verortet werden kann, oder seien es begriffliche Präzisierungen in der Beschreibung von technikphilosophischen Teilgebieten. Die folgende Aussage von Goodman zu seiner Methode in *Ways of Worldmaking* fasst auch die Vorgehensweise des zweiten Hauptteils elegant zusammen.

This book does not run a straight course from beginning to end. It hunts; [...] and it counts not the kill but what is learned of the territory explored.

(Goodman 1979)

Insgesamt erreicht der zweite Hauptteil jedoch zumindest ein Etappenziel und endet mit dem Vorschlag der Begriffe der Neugier, Vorstruktur und Welttechnik zur Beschreibung der systematisch neuen Aspekte maschinellen Lernens. In Anstrebung dieses Ziels versucht der zweite Hauptteil zu klären, was unter einer technisch unterstützten Erschlossenheit von Welt verstanden werden kann und welche Aspekte dieses typischerweise exklusiv dem Menschen – allenfalls mitunter auch Tieren oder **Organismen** – zugesprochenen Begriffes in Bezug auf MLA verwendbar sind. Diese Klärung erfolgt parallel zur begrifflichen Fixierung des maschinellen Lernens. Begonnen wird diese Aufarbeitung über eine Diskussion, der in *Sein und Zeit* (Heidegger 1927) ausgearbeiteten Daseinsanalytik. Die Übertragung der Konzepte des In-Seins, des Verstehensbegriffes und anderer Elemente der Heideggerschen Begriffswelt auf das maschinelle Lernen ist jedoch nicht angestrebt. Der gewonnene Mehrwert wird hier insbesondere die Nutzbarmachung eines zumindest neutral konnotierten Begriffes von Neugier als Problemittel für das sein, was MLA aus technikphilosophischer Sicht interessant macht.

Der zweite Hauptteil schließt ab mit einer **Partitionierung**¹² des maschinellen Lernens in zwei Ausprägungen – neugieriges und zielorientiertes Lernen – und der Verortung der beiden Ausprägungen relativ zu den Begriffen klassischer und transklassischer Technik. Abgeleitet aus den Ergebnissen wird der Begriff der Welttechnik motiviert und entworfen und schließlich wird der Zusammenhang von Welttechnik und **Nichtwissen** dargestellt.

1.3.3. Interdisziplinarität

Eine Reihe von Denkansätzen, wie etwa die Rede von **konvergierenden Technologien** oder verschiedene Begriffsbildungen zu **transklassischer Technik**, argumentieren abstrakt und damit technikfern, dass in gewissen Sektoren neue Techniken entstehen oder vorliegen, deren Analyse komplett neue Technikkonzepte erfordert. Die aus dieser Argumentationsweise entstehende Skepsis und die Grundmotivation dieser Arbeit liegen darin, techniknah prüfen zu wollen, was maschinelles Lernen ausmacht und darauf aufbauend festzustellen, ob sich

¹² Eine Unterteilung des vollständigen Feldes in überschneidungsfreie Teilfelder.

spezielle Teilbereiche dieser Form von Technik tatsächlich nicht mit etablierten Technikkonzepten fassen lassen.

Der daraus resultierende Anspruch dieser Analyse liegt darin, ein interdisziplinäres Verständnis der betrachteten Technik zu vermitteln und dabei mögliche technikphilosophische Fragestellungen aufzuzeigen. Weder sollen reine Falluntersuchungen durchgeführt noch die zugrunde liegenden Algorithmen aus den Augen verloren werden. Die zur Entwicklung eines technikenahen Verständnisses notwendige, interdisziplinäre **Grundlagenforschung** soll im ersten Hauptteil dieser Analyse vorangebracht werden. Das Ziel des zweiten Teils liegt darin, einen interdisziplinären Austausch zwischen Informatik und Philosophie zu ermöglichen, indem eine Ausgangsbasis für die technikphilosophische Verortung des maschinellen Lernens geschaffen wird. Schon für die weiterführende Diskussion des im zweiten Hauptteil angedachten Begriffes der Welttechnik ist ein erweitertes Verständnis für maschinelles Lernen hilfreich, wenn nicht notwendig. Aus diesem Grund wird die Übersetzung der Vokabeln des maschinellen Lernens in die Begriffe der Technikphilosophie etwas ausführlicher durchgeführt als es – rückblickend vom Ergebnis, das der zweite Hauptteil anstrebt – unbedingt nötig wäre. Auch die Frage, ob und inwiefern maschinelles Lernen Technik ist, kann gerade nicht nur rein theoretisch gestellt werden, sondern muss im Hinblick auf implementierte Algorithmen formuliert werden. Erst wenn ausreichend viele Aspekte der konkreten Technik interdisziplinär verstanden und beschrieben wurden, kann gehofft werden, mit einer abstrakten oder begrifflichen Betrachtung einen Mehrwert zu schaffen. Weiterhin ist die resultierende Diskussion nicht notwendigerweise beschränkt auf maschinelles Lernen. Die Auflösung des Begriffs der Selbstorganisation in der Informatik lässt sich auf andere Technikbereiche übertragen – wenn auch nur in Zusammenhang mit detailliertem Fachwissen in der jeweiligen Disziplin.

Das Alleinstellungsmerkmal dieser Arbeit und der Inhalt des ersten Hauptteils können zusammengefasst werden als geisteswissenschaftliche Erarbeitung einer Perspektive auf die technischen Grundlagen eines neuen Gebietes der Informatik. Die Diskussion des zweiten Hauptteils ist ebenfalls sehr technikenah gehalten und abstrahiert im Gegensatz zu theoretischeren Projekten nur einmal von der faktisch implementierbaren Technik¹³. Die resultierende Darstellung kann und soll als Grundlage für Überlegungen zu den genannten Anknüpfungspunkten und zu völlig anderen als den vorgestellten Stoßrichtungen dienen.

¹³ Im Rahmen der Suche nach Problembegriffen zur Eröffnung einer technikphilosophischen Diskussion des maschinellen Lernens in Abschnitt 3.4.

2. Erster Hauptteil: Der Blick der Informatik auf maschinelles Lernen

2.1. Betrachtungsebenen

Ziel des ersten Hauptteils ist es, eine interdisziplinär verständliche techniknahe Beschreibung maschinellen Lernens als konkretem Teil der Informatik zu erstellen. Zu diesem Zweck wird eine geisteswissenschaftliche Innenansicht der Perspektive der Informatik erstellt. Einleitend hierzu noch einmal die bereits genannte Definition maschinellen Lernens:

Verfahren des machine learning sind die Grundlage von Programmsystemen, die aus „Erfahrung“ lernen, also neues Tatsachen- und Regelwissen gewinnen oder Priorisierungen adaptieren können. Sie sind u.a. auch für die Entdeckung zweckbestimmt relevanter Beziehungen in großen Datenmengen („Data mining“) von großer Bedeutung.

(Görz et al. 2003, S. 13)

Vor einer Aufarbeitung des maschinellen Lernens muss entschieden werden, auf welcher Abstraktionsebene das Gebiet untersucht werden soll. Auf der *höchsten* Ebene wären Themen der **theoretischen Informatik** wie Berechenbarkeitstheorie oder Komplexitätstheorie zu behandeln. Entsprechend der dieser Arbeit zugrunde liegenden Idee soll keine Diskussion von Metaperspektiven auf die Informatik oder auf Algorithmen stattfinden. Die nachfolgenden Betrachtungen beschreiben auf einer *zweiten*, weniger abstrakten Ebene die Auflösung des maschinellen Lernens in Klassen von Algorithmen und basieren auf Standardwerken der Informatik (Mitchell 1997; Görz et al. 2003; Russell et al. 2007; Alpaydin 2008; Brause 2010; Burkhard 2010) und der Wikipedia¹⁴ (Wikipedia 2001). Die Auswahl und Hierarchisierung der **Lernstrategien** stellt jedoch eine genuin eigene Klassifikation dar, die im Hinblick auf die interdisziplinäre Relevanz erstellt werden wird. In der Literatur liegen viele unterschiedliche Varianten der Bestimmung von Lernstrategien vor¹⁵, demnach stimmt die in Abschnitt 2.3 vorgenommene Klassifikation mit manchen dieser Varianten zu größeren Teilen und mit manchen nur sehr eingeschränkt überein. Die *dritte* mögliche Abstraktionsebene entspricht der Verwendung von **Pseudocode** zur Diskussion von Algorithmen. Ein Beispiel für einen in Pseudocode geschriebenen Algorithmus ist das folgende Pseudoprogramm.

¹⁴ Grundlage sind verschiedene Artikel der deutsch-, englisch- und in geringem Maße der französischsprachigen Version der Wikipedia. Zitiert wird im Weiteren jedoch ausschließlich aus der deutschsprachigen Version.

¹⁵ Sogar die englische, deutsche und französische Version der Wikipedia unterscheiden sich stark. Die deutsche Wikipedia führt etwa zunächst keine Unterscheidung zwischen Klassen maschinell lernender Algorithmen durch, sondern ordnet diese nach überwachtem, unüberwachtem und bestärkendem Lernen.

```
Programm: Essen_kochen
Variablen: Appetit, Gericht, Einkaufsliste
■ Gericht=Gericht_auswählen(Appetit)
■ Einkaufsliste = Liste_erstellen(Gericht)
■ WIEDERHOLE
    Einkaufen
■ BIS Einkaufsliste = LEER
■ Kochen(Gericht)
■ ENDE
```

Abbildung 1: Beispiel für Pseudocode

Solcher Pseudocode kann verwendet werden, um über ein Programm zu sprechen, ohne die Syntax einer speziellen Programmiersprache verwenden zu müssen. Im Weiteren wird es um die Veranschaulichung von Algorithmen gehen und gelegentlich wird Pseudocode verwendet werden, um diese Algorithmen zu veranschaulichen. Diese Abstraktionsebene ist passend für die interdisziplinäre Diskussion maschinellen Lernens, allerdings müssen die nächsten beiden Ebenen zumindest noch mitgedacht werden. Auf der *vierten* Abstraktionsebene werden konkrete Algorithmen in einer speziellen Programmiersprache betrachtet. Ein Stück Pseudocode kann sehr unterschiedlich in konkrete Algorithmen umgeformt werden. Einige im Weiteren genannte Beispiele und Leistungskennzahlen werden sich auf konkrete Algorithmen beziehen, die Diskussion wird jedoch auf der dritten oder auf der zweiten Abstraktionsebene stattfinden und lediglich zwischen einzelnen Algorithmenklassen unterscheiden. Dieses Auflösungsvermögen ist bereits eine vergleichsweise starke Forderung, da in der Technikphilosophie üblicherweise nicht zwischen autoadaptiven und nicht-autoadaptiven Algorithmen unterschieden wird. Ein gewisses Verständnis der konkreten Algorithmen ist dennoch hilfreich, um ungefähr zu überblicken welche Leistungsfähigkeit verschiedene Formen maschinellen Lernens in der Praxis aufweisen. Die über die Betrachtung der konkreten Algorithmen hinausgehende *fünfte* Abstraktionsebene schließlich ist die Implementierung eines Algorithmus in einem physischen Objekt. Häufig werden die vierte und die fünfte Ebene nicht unterschieden, da die Programmierung in einer speziellen Programmiersprache nur theoretisch unabhängig von der Implementierung in einem elektronischen Bauteil erfolgen kann. In der Praxis des Programmierens werden ständig programmexterne Ressourcen beziehungsweise Funktionen benötigt und viele Details der Syntax werden üblicherweise von der Programmierungsumgebung automatisch beachtet. Dennoch ist es theoretisch möglich, die beiden Ebenen zu trennen, etwa kann man einen konkreten Algorithmus auch auf einem Stück Papier ausdrucken. Die Unterscheidung der vierten und fünften Ebene soll an dieser Stelle lediglich explizit machen, dass auch MLA immer schon eine fünfte Ebene aufweisen und dass deswegen nicht pauschal von maschinell lernenden *Algorithmen* gesprochen werden sollte¹⁶. Wenn die Hardware mitgedacht wird, können viele Missverständnisse vermieden werden, insbesondere solche über den Zusammenhang zwischen einer Nutzeroberfläche und einem MLA, sowie die Schwierigkeiten bei der Feststellung, ob ein MLA im jeweils aktuellen Zustand noch lernt oder ob der Lernvorgang schon abgeschlossen

¹⁶ Ein Algorithmus selbst kann nicht auf Reize reagieren, da er hierfür Eingabedaten benötigt, die erst dann auftreten können, wenn er in einem elektronischen Bauteil implementiert wurde.

ist. So wie ein Künstler nicht mit seinem Kunstwerk identisch ist, lassen sich auch MLA von ihren Strukturvorschlägen unterscheiden und dies gelingt besser, wenn die Hardware mitgedacht wird. Die Abkürzung MLA bezieht sich dementsprechend gezielt auf Artefakte, auch wenn die genaue Ausprägung des Artefaktes im Weiteren die meiste Zeit nicht von besonderer Bedeutung ist.

Eine Diskussion, bei der die konkrete Form des Artefaktes zentral ist, ist diejenige um das **Ubiquitous Computing** oder **UbiComp** beziehungsweise die **Ambient Intelligence**¹⁷. Das UbiComp wird im Weiteren nicht im Fokus stehen, allerdings erscheint es zunächst wie eine Betrachtung des maschinellen Lernens auf der fünften Abstraktionsebene. Zur Vermeidung einer falschen Erwartung an die nachfolgende Diskussion und zum besseren Verständnis der fünften Abstraktionsebene soll daher kurz dargestellt werden, was unter Ubiquitous Computing verstanden werden kann. Im UbiComp geht es insbesondere darum, **smarte** Artefakte zu konzeptionieren und zu konstruieren, das heißt Artefakte, die einen Nutzerwunsch antizipieren und erfüllen, bevor der Nutzer den Wunsch äußern kann oder muss. Die Artefakte treten dabei selbst nicht in Erscheinung, sondern sind ein unaufdringlicher Teil der Umwelt des Nutzers. Ein Beispiel ist die automatische Aktivierung der Beleuchtung, wenn man einen Raum betritt. Viele Artefakte des UbiComp basieren auf maschinellern Lernen, da ein Artefakt, das mittels eines Autoadaptionprozesses aus den Aktionen des Nutzers lernt, besonders gut geeignet, ist zukünftige Aktionen zu antizipieren und smart zu agieren, bevor der Nutzer aktiv werden muss. Auch wenn das konkrete Artefakt im UbiComp nicht lernt, ist in dessen Erstellung häufig maschinelles Lernen eingeflossen, da in vielen Fällen automatisch das Verhalten von vielen anderen Nutzern beobachtet wurde, um ein Artefakt konstruieren zu können, das besonders gute Chancen hat, unauffällig im Hintergrund arbeiten zu können. Maschinelles Lernen wird daher häufig mitgedacht, wenn vom Bereich des UbiComp die Rede ist. Einerseits ist es jedoch in der diesbezüglichen Diskussion nicht immer zentral, woher die Artefakte die Regeln für ihre Vorgehensweise erhalten haben und andererseits muss ein MLA nicht zwangsweise unspürbar im Hintergrund agieren. UbiComp ist entsprechend zwar eine zentrale, aber nicht die einzige Anwendungsmöglichkeit für maschinelles Lernen. Ein Anknüpfungspunkt an die Verbindung zwischen UbiComp und der im zweiten Hauptteil entworfenen Welttechnik besteht in dem Konzept des **Flows** (Hassenzahl 2004, S. 16). Der Flow beschreibt die selbstvergessene Nutzung eines Artefaktes und stellt gegebenenfalls ein verwandtes Phänomen zu UbiComp und Welttechnik dar.

¹⁷ Beide Begriffe werden aktuell innerhalb der Informatik annähernd synonym verwendet, wobei die Bezeichnung Ubiquitous Computing häufiger anzutreffen ist.

2.2. Unterschiede zwischen lernenden Algorithmen

Nachdem die Auflösung der Analyse festgelegt wurde, ist der nächste Schritt einerseits zu bestimmen, nach welchen Kriterien maschinell lernende Algorithmen unterschieden und andererseits, wie sie zu Algorithmenklassen – den **Lern-** beziehungsweise **Autoadaptionsstrategien** – zusammengefasst werden können. Im Folgenden sollen zuerst die in der Informatik gebräuchlichen Unterscheidungsmerkmale zwischen lernenden Algorithmen dargestellt werden.

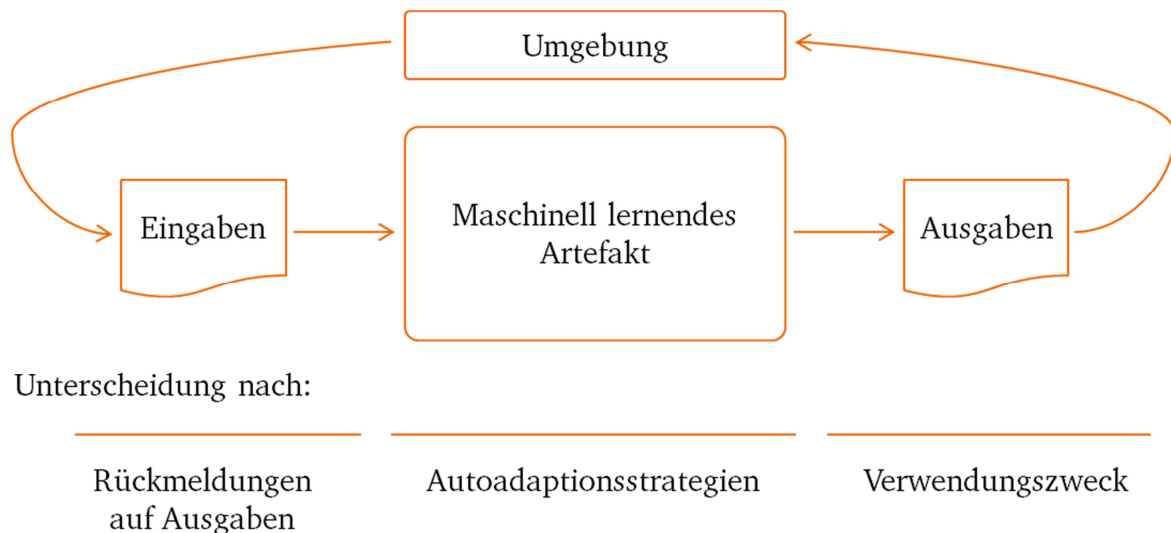


Abbildung 2: Überblick der Möglichkeiten zur Unterscheidung von MLA

Diese Unterscheidungen finden sich in ähnlicher Form in jedem der eingangs genannten Standardwerke zu maschinellem Lernen. Die Unterscheidungen betreffen den gesamten Bereich des maschinellen Lernens und werden im Folgenden als Einstieg genutzt, um die für das maschinelle Lernen relevanten formalen Grundbegriffe der Informatik einzuführen. Die Lernstrategien fallen mitunter etwas spezieller aus und bringen jeweils eigene, klassenspezifische Vokabeln mit.

2.2.1. Unterscheidung gemäß erhaltener Rückmeldungen

Das erste Unterscheidungskriterium für maschinell lernende Algorithmen und die korrespondierenden MLA basiert auf der Beobachtung, wie verschiedene Algorithmen im Rahmen ihres Autoadaptionsprozesses eine Rückmeldung über die Qualität der ausgegebenen Strukturen erhalten. Die Eingabedaten, die ein lernender Algorithmus während des Autoadaptionsprozesses erhält, werden in der Informatik als **Trainingsinstanzen** bezeichnet, um sie von den **Testinstanzen** abzugrenzen, die gemäß ihres Namens verwendet werden um ein MLA nach Abschluss des Autoadaptionsprozesses zu testen. Im zweiten Hauptteil wird meist an Stelle von Trainingsinstanzen die Rede von **Rohdaten** sein, wenn Daten gemeint sind, die ein Nutzer einem MLA übergibt und die einen Autoadaptionsprozess initiieren sollen, dessen Ergebnis zumindest teilweise offen ist. Demgegenüber wird von **Trainingsdaten** gesprochen, wenn das Ergebnis des Autoadaptionsprozesses im Vorfeld bereits detailliert festgelegt wurde. Diese Begriffsverschiebung motiviert sich zum einen daraus, dass der Begriff der Trainingsinstanzen den Eindruck erweckt, dass die Daten eine Struktur instanziierten und zum anderen wird impliziert, dass Algorithmen trainiert werden können. Ebenso leidet das Verständnis der Diskussion der Unterschiede zwischen lernenden Algorithmen, wenn

Autoadaptionsergebnisse pauschal als von den Eingabedaten instanziiert gedacht werden. Die Frage, ob Rohdaten etwas instanziiieren, wird im zweiten Hauptteil noch ausführlicher diskutiert, aber schon die Diskussion, ob autoadaptive Algorithmen im menschlichen Sinn lernen, soll hier vermieden werden und das gilt ebenso für die Diskussion, ob Algorithmen trainieren oder trainiert werden. Im zweiten Hauptteil wird aus diesem Grund neben dem Begriff der Trainingsinstanzen auch der präzisere Begriff der Trainingsdaten soweit wie möglich vermieden. Im Abschnitt zur Klassifikation von Lernstrategien wird der Begriff der Trainingsinstanzen allerdings dennoch Verwendung finden. Die reflektierte Rede von Trainingsinstanzen hilft in diesem Fall dabei, die Perspektive der Informatik auf das maschinelle Lernen präziser darstellen zu können, denn bei der Klassifikation von Lernstrategien wird es weniger um formale Kriterien auf der Betrachtungsebene der Algorithmen und mehr um Ideen und Konzepte auf der Betrachtungsebene von Pseudocode gehen. Der Trennung zwischen Trainingsdaten und Rohdaten liegt somit die Unterscheidung zwischen unterschiedlichen Formen der Rückmeldung auf Ausgabedaten und damit eines der wesentlichsten Kriterien zur Differenzierung von maschinell lernenden Algorithmen zugrunde. Die unterschiedlichen denkbaren Formen der Rückmeldung lassen die Differenzierung von drei Varianten maschinellen Lernens zu. Diese drei Varianten werden als überwachtes Lernen, unüberwachtes Lernen und bestärkendes Lernen bezeichnet.

In fast allen Fällen nimmt ein MLA Eingabedaten auf und gibt Ausgabedaten aus – nur in Sonderfällen besitzt ein MLA keine formale Ausgabe, so dass die sich verändernde Struktur des MLA selbst analysiert werden muss beziehungsweise kann. In Konsequenz möchte ein Nutzer oder Entwickler ein MLA häufig dazu bringen, dass der Eingabe-Ausgabe-Zusammenhang eine bestimmte Funktion abbildet. Zu diesem Zweck kann er Eingabedaten zur Verfügung stellen, bei denen die korrekte beziehungsweise gewünschte Ausgabe bekannt ist und mit angegeben werden kann. In diesem Fall spricht man von **überwachtem Lernen**. Bei überwachtem Lernen erhält das MLA während des Autoadaptionsprozesses Paare von Ein- und Ausgabewerten und das Ziel liegt darin eine Struktur zu erzeugen, die nach Abschluss des Lernprozesses auf Erhalt des Eingabedaten-Anteils eines Trainingsdatums den Ausgabedaten-Anteil ebenjenes Datenpunktes zurückgibt. Zur Veranschaulichung soll eine automatische Zahlenerkennung betrachtet werden. Es soll angenommen werden, dass dazu im Vorfeld eine Reihe von Bildern von Zahlen digital erfasst und vom Nutzer mit den entsprechenden numerischen Werten versehen wurde. Das MLA bekommt während des Autoadaptionsprozesses als Eingabe Datenpaare, deren Teile die Eingabe und die Ausgabe für das zu erstellende Lernergebnis darstellen. Ein Beispiel wäre die Übergabe des ersten Zahlenbildes zusammen mit der Aussage, dass dort eine Acht zu sehen ist. Das MLA könnte anschließend die Regel „wenn 20 bis 21 Prozent des Bildes schwarz sind, ist die Ausgabe eine acht“ erstellen. Wenn die Bilder der zu kategorisierenden Zahlen alle auf digitalen Darstellungen derselben Schriftart basieren, kann das MLA zehn Regeln dieser Art erstellen und hat die **Zielfunktion** erlernt. Überwachtes Lernen wird häufig eingesetzt, wenn über die Struktur der Rohdaten schon Vorwissen besteht oder das Vorliegen gewisser Strukturen zumindest vermutet wird. Wenn etwa die Rohdaten aus Messwerten bestehen, die einem bestimmten Muster entsprechen, wird dieses Muster als die **Verteilung** der Rohdaten bezeichnet. Die Messwerte bei Beobachtung eines Würfelwurfes können beispielsweise die Werte von eins bis sechs sein und bei symmetrischen Würfeln treten alle Würfelergebnisse mit ungefähr der gleichen Wahrscheinlichkeit auf – die Ergebnisse sind in diesem Fall **gleichverteilt**. Nachfolgend zu Veranschaulichung die

Verteilung der Messwerte für einen asymmetrischen Würfel, der dies nicht ganz erfüllt und häufiger die Sechs und dafür weniger häufig die Eins zeigt.

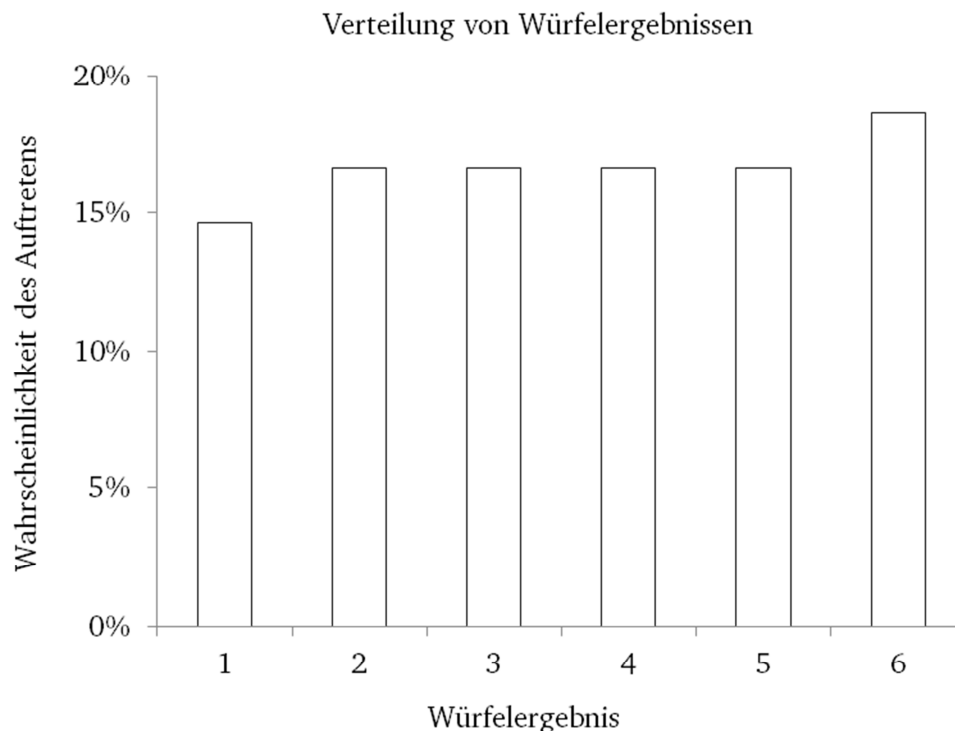


Abbildung 3: Beispielverteilung von Würfelergebnissen eines asymmetrischen Würfels

Mit Hilfe eines MLA kann versucht werden, die Details der Verteilung der Würfelergebnisse eines speziellen, asymmetrischen Würfels zu bestimmen. In diesem Fall würde angenommen werden, dass die Würfelergebnisse unabhängig voneinander und zumindest ungefähr gleichverteilt sind. Wenn weiter über die Verteilung der Rohdaten im Vorfeld bekannt ist, dass jedes Würfelergebnis mit einer unbekannten aber festen Wahrscheinlichkeit auftritt und dass nur sechs Ergebnisse möglich sind, bedeutet das, ein MLA muss sechs **Parameter** bestimmen¹⁸, um die Details der Verteilung der Würfelergebnisse aufzuzeigen. In der Anwendungspraxis maschinellen Lernens wird sehr häufig im Vorfeld ein Modell erstellt, um darauf aufbauend die für den Autoadaptionsprozess relevanten Parameter zu bestimmen. Ein MLA erlernt anschließend diejenigen Werte für diese Parameter mit Hilfe derer das Modell den Rohdaten am besten entspricht. Überwachtes Lernen bietet sich besonders dann an, wenn ausreichend Messwerte vorliegen und in erster Linie die Parameter des ausgewählten Modells optimiert werden sollen. Im Falle des Würfelwurfes ist die zugrunde liegende Struktur sehr einfach und ein MLA würde nur wenig Mehrwert bieten, da die Auftretenswahrscheinlichkeiten und damit die Parameter der Verteilung auch unkompliziert vom Nutzer selbst auf Basis der Würfelergebnisse errechnet werden könnten.

Das überwachte Lernen setzt die Verfügbarkeit von bekannten Trainingsdaten im Sinne von Eingabe-Ausgabe-Paaren voraus. Sollten solche Trainingsdaten nicht zur Verfügung stehen, kann **unüberwachtes Lernen** zum Einsatz kommen. Hier sind nur Eingabewerte in Form von Rohdaten gegeben und es sind Regeln und Modelle gesucht, nach denen die Eingabedaten

¹⁸ Zu erlernen sind eigentlich nur fünf Parameter, da die Wahrscheinlichkeit für das sechste Ergebnis berechenbar ist, wenn die ersten fünf Wahrscheinlichkeiten bekannt sind.

strukturiert werden können. Ein Beispiel für die Erstellung solcher Modelle ist die **Clusteranalyse**. Hierbei sollen noch unbekannte Rohdaten in Gruppen von ähnlichen Daten eingeteilt werden. Dies sollte nicht verwechselt werden mit der **Klassifizierung** von Daten, bei der die Klassen bereits im Vorfeld festgelegt wurden und die Rohdaten den Klassen zugeordnet werden sollen. Die Clusteranalyse erzeugt **Cluster** genannte Klassen von Rohdaten nach einem vorgegebenen oder erlernten **Ähnlichkeitsmaßstab**. Ein verwandtes Beispiel für unüberwachtes Lernen ist die Suche nach **Assoziationsregeln** in Rohdaten, das heißt nach Aussagen, die für große Teile der Rohdaten zutreffend sind. Ein Beispiel für eine Assoziationsregel zu einer fehlenden Assoziation wäre die Aussage, dass die Wahrscheinlichkeit eine Zwei zu würfeln nicht davon abhängt, ob im vorherigen Versuch eine Eins oder eine Vier gewürfelt wurde. Ein anderes Beispiel ist die Suche nach möglichen Kaufempfehlungen, die sich darauf beziehen, dass bestimmte Produkte häufig zusammen gekauft werden.

Eine häufig eingesetzte und illustrative Mischform maschinellen Lernens verbindet menschliches Vorwissen, unüberwachtes und überwachtes Lernen. Diese Mischform beginnt damit, dass vom Nutzer zu einer Menge von Rohdaten mehrere Modelle bestimmt werden, die jeweils Teile der Struktur der Rohdaten abbilden. Anschließend werden die Rohdaten unüberwacht durch ein MLA in Cluster unterteilt und es werden für jedes Cluster überwacht unterschiedliche Modelle parametrisiert, um optimale **lokale Modelle** zu bestimmen. In diesem Fall müssen nutzerseitig für jedes Cluster Trainingsdaten beschafft werden, typischerweise über die Durchführung von Messungen oder über eine komplexe, das heißt zeitraubende mathematische Berechnung. Diese Mischform maschinellen Lernens kann dazu dienen, die Anzahl kostspieliger Messungen oder den Zeitaufwand der notwendigen Berechnungsprozesse zu reduzieren, indem nur lokale Aussagen angestrebt werden und kein einheitliches Modell für alle Rohdaten gesucht wird. Die nach Abschluss des Autoadaptionsprozesses entstandene Struktur kann beim Auftreten neuer Rohdaten typischerweise sehr schnell ausgewertet werden. Ein Beispiel ist die Suche nach einem optimalen Fahrverhalten für einen Autopiloten für Automobile. Hierbei kann das MLA zuerst die Strecke in Geraden und verschiedene Arten von Kurven unterteilen und anschließend für die Teilstücke ein optimales Beschleunigungs- und Lenkverhalten erlernen – etwa indem auf einem ebenen und großen Asphaltstück nur Kurven gefahren werden.

Die dritte Variante neben überwachtem und unüberwachtem Lernen ist das **bestärkende Lernen**. In diesem Fall führen MLA **Sequenzen** von Aktionen durch, deren Einzelschritte jeweils keine Bewertung durch den Nutzer herbeiführen und deren Länge variabel ist. Nur an ausgewählten Zwischenschritten beziehungsweise Punkten der Sequenz und nach Abschluss der Sequenz erhält das MLA eine Rückmeldung auf seine Ausgabe. Diese Vorgehensweise ist an einem Beispiel schnell verständlich. Angenommen, ein MLA soll erlernen Schach zu spielen, dann ist meist nicht bekannt, ob ein spezieller Zug besser oder schlechter ist als ein anderer möglicher Zug. Gleichzeitig können recht einfach die Bedingungen festgelegt werden, unter denen eine Zugfolge des MLA mit einem Sieg oder einer Niederlage beendet ist. Eine solche Situation erfüllt genau die beschriebenen Voraussetzungen für bestärkendes Lernen. Bestärkendes Lernen eignet sich dementsprechend gut für dynamische Umgebungen, da nur das Ziel des Autoadaptionsvorgangs vorgegeben wird. Die Durchführung bestärkenden Lernens setzt häufig auf **Brute-Force**. Diese Vorgehensweise simuliert schlicht alle oder doch möglichst viele der möglichen Sequenzen von Aktionen, um dann diejenige Sequenz mit der besten Gesamtbewertung auszuwählen. Diese Methode stößt jedoch schon beim Schachspiel

an ihre Grenzen, da mehr als 10^{40} Stellungen¹⁹ möglich sind (Shannon 1949). Eine weniger aufwändige Variante der Brute-Force-Methode betrachtet eine große, aber beschränkte Anzahl von Sequenzen und schätzt deren Bewertung ab, um anschließend aus diesen Sequenzen eine neue Sequenz zu erstellen, die die Bewertung optimiert. Bestärkendes Lernen mit einer Sequenz der Länge eins schließlich entspricht formal genau dem überwachten Lernen. Bestärkendes Lernen kann auch in einigen weiteren Fällen als eine abgeschwächte Form des überwachten Lernens betrachtet werden, insbesondere bei kurzen Sequenzen mit vorgegebener Länge. Allerdings ist schon am Beispiel des Schachspiels ersichtlich, dass zwischen Anfang und Ende der Sequenz eine sehr große Vielzahl von Zügen möglich ist und dass dies nicht sehr gut der Idee des überwachten Lernen, ein Modell zu parametrisieren, entspricht.

2.2.2. Unterscheidung nach Suchstrategien

Eine andere Möglichkeit die Algorithmen des maschinellen Lernens zu klassifizieren, besteht darin, die eingesetzte Suchstrategie zu betrachten. Diese Betrachtung ist in gewisser Hinsicht verwandt mit der vorherigen Klassifizierung, da für die Analyse der Suchstrategien der Fokus unter anderem darauf gelegt wird, was gesucht werden kann und damit primär auf die Ausgaben des MLA – im Gegensatz zur obigen Fokussierung auf die Eingabedaten. Die Menge aller möglichen Ausgaben eines MLA bildet deren **Lösungs-, Such-, oder Hypothesenraum**. Die drei Formulierungen werden in der Informatik meist synonym verwendet, können im Einzelfall jedoch auch betonen, dass das MLA ein klar umrissenes Problem lösen, einen Raum möglicher Ausgaben durchsuchen oder mögliche Konzepte und Modelle zu Rohdaten vergleichen oder suchen soll. Der Begriff des Hypothesenraumes scheint die Implikationen eines die Hypothese formulierenden Agenten mitzuführen. Das MLA wird in diesen Fällen jedoch nicht als konzeptbildend gedacht. Stattdessen werden sehr starke Vorstrukturierungen vorgenommen, die Konzepte abbilden und das Artefakt vergleicht oder optimiert diese Konzepte auf Basis von Eingabedaten. Weiterhin wird von einem Hypothesenraum aus Sicht vieler Entwickler oder Nutzer genau dann gesprochen, wenn systematische Fehler von Algorithmen aufgedeckt werden sollen. Die Ausgabe eines MLA – sei es ein Element des Such-, Lösungs- oder Hypothesenraumes – wird im Weiteren, wie in Abschnitt 1.2 beschrieben, als **Strukturvorschlag** bezeichnet. Einen Strukturvorschlag kann schon die Unterteilung einer Speisekarte in vegetarische und nicht-vegetarische Gerichte darstellen und die Ergebnisse der Autoadaptionsprozesse vieler MLA bestehen auch genau in solchen Strukturen geringer Komplexität²⁰. Ein Beispiel aus der Praxis besteht darin, die Frage was Paris zu Paris macht, zu beantworten, indem eine automatische Zusammenstellung der das Stadtbild prägendsten visuellen Elemente erstellt wird (Doersch et al. 2012). Im Beispiel der Schach erlernenden MLA würde sich der Lösungsraum aus der Menge aller denkbaren Sequenzen zusammensetzen. Ein Entwickler könnte in diesem Fall den Lösungsraum beschränken, indem nur ein gewisser Fundus an Eröffnungen gespielt werden darf und indem für das Endspiel eine Datenbank hinzugezogen werden muss, die eine Vorgehensweise vorschreibt. Solche Vorgaben werden sehr häufig gemacht um den Autoadaptionsprozess schneller zu einem erfolgreichen Ende zu bringen.

¹⁹ Die Anzahl der Atome im Körper von 10 Milliarden Menschen kann auf 10^{34} Potenz geschätzt werden (Bauer et al. 1999). Die Anzahl der möglichen Stellungen ist entsprechend unvorstellbar groß.

²⁰ Woraus nicht gefolgert werden sollte, dass der Weg hin zu Strukturvorschlägen geringer Komplexität ebenfalls trivial ist. Ein gut verständliches und sofort nützliches Lernergebnis muss nicht bereits im Vorfeld offensichtlich gewesen sein.

Wenn der Suchraum eines MLA nun aus Elementen beziehungsweise Strukturen besteht, die sich sinnvoll anordnen lassen, kann eine **geordnete Suche** durchgeführt werden. Wenn etwa Passwörter gefunden werden sollen, kann dies mittels maschinellen Lernens versucht werden. Der Ausgaberaum wäre in diesem Fall alphanumerisch sortierbar und der Strukturvorschlag könnte nach Eingabe einiger Informationen über den Passwortinhaber die Kennwörter aller „ähnlichen“ Nutzer nach der Häufigkeit von deren Auftreten angeordnet ausgeben. In diesem Fall wäre die Aktualisierung der Ausgabereihenfolge auf Basis des Erfolges der Ausgaben ein möglicher Autoadaptionsprozess.

Eine alternative Suchstrategie besteht darin, einzelne Parameter, die die Entstehung von Strukturvorschlägen oder Lösungen beeinflussen, zu manipulieren, die entstehende Lösung zu bewerten und die Bewertungen zu vergleichen. Anschließend werden die betrachteten Parameter so angepasst, dass die resultierende Lösung in einer gewissen Hinsicht optimal ist. Ein abstraktes, aber anschauliches Beispiel hierfür ist der Versuch, in Modellrechnungen die Verkehrssicherheit von Pkw zu betrachten und dabei als Parameter die Bremskraft und die Genauigkeit der Tankanzeige zu betrachten. In diesem Beispiel wird sich voraussichtlich zunächst herausstellen, dass eine Manipulation des Parameters der Genauigkeit der Tankanzeige die Verkehrssicherheit nicht signifikant beeinflusst und somit verworfen werden kann. Anschließend wird festgestellt werden, dass die Bremskraft tatsächlich einen systematischen Einfluss auf die Sicherheit hat und ein höherer Wert dieses Parameters sehr häufig vorteilhaft ist. Das bedeutet, der resultierende Vorschlag wäre, den Parameter der Bremskraft zu erhöhen. Diese Form der Suche wird als **Gradientensuche** bezeichnet und identifiziert die optimale Veränderung gegebener Parameter. Eine Gradientensuche, bei der wie im genannten Beispiel nur ein Parameter verändert werden kann oder soll, lässt sich meist auch als eine geordnete Suche darstellen. Die Suche im Beispiel etwa ordnet Pkw systematisch nach deren Bremskraft an und prüft anschließend im Rahmen des Autoadaptionsprozesses die Auswirkung einer veränderten Bremskraft auf die Verkehrssicherheit. Die Suchergebnisse werden in diesem Fall als durch die Größe des beeinflussbaren Parameters geordnet gedacht. Der Fokus einer Gradientensuche und einer geordneten Suche ist dennoch unterschiedlich, da im ersten Fall mathematisch die optimalen Parameter zur Erstellung einer Lösung gesucht werden und im zweiten Fall die Reihenfolge festgelegt wird, in der mögliche Lösungen betrachtet werden sollen. Die dritte wesentliche Ausprägung neben der geordneten Suche und der Gradientensuche ist die **stochastische Suche**. Bei der stochastischen Suche soll eine **Verteilung** gefunden werden, die optimal eine Menge von Rohdaten modelliert. Die Rohdaten werden als eine **Stichprobe** für die zugrunde liegende Verteilung interpretiert und gemäß der folgenden Illustration wird eine Suchstrategie ausgewählt.

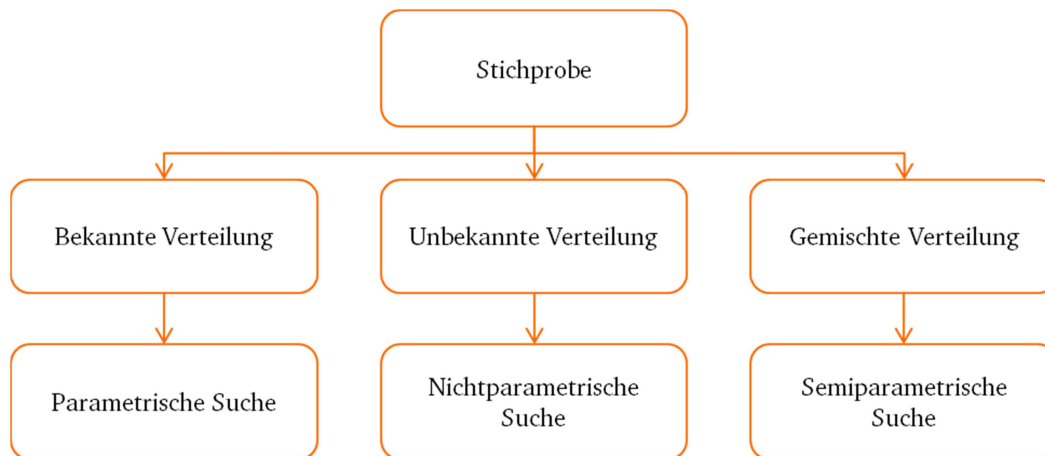


Abbildung 4: Überblick der Möglichkeiten stochastischer Suche

Parametrische Suchen und nichtparametrische Suchen wurden bereits in der Klassifizierung von Algorithmen über deren Eingabedaten kurz dargestellt. **Semiparametrische Suchen** stellen verschiedene Mischformen dar, etwa wenn lokale Modelle erstellt werden, die zum Teil mit parametrischen und zum Teil mit nichtparametrischen Ansätzen bearbeitet werden.

2.2.3. Unterscheidung nach Verwendungszweck

Mitunter wird das maschinelle Lernen auch auf Basis des Verwendungszwecks in Teilbereiche unterteilt. Ein Beispiel ist die Unterscheidung zwischen den Typen des parametrischen, semiparametrischen und nichtparametrischen Lernens. Diese Unterscheidung verortet geordnete Suchen und Gradientensuchen je nach Einzelfall unter einem der drei Typen. Meist wird dabei davon ausgegangen, dass nichtparametrisches Lernen in erster Linie dann eingesetzt werden kann, wenn ein starker und stabiler Zusammenhang zwischen Eingaben und Ausgaben besteht und Änderungen dieses Zusammenhangs nur langsam auftreten. Solch ein Zusammenhang wird bei Gradientenverfahren benötigt, da diese das Vorliegen eines direkten Zusammenhangs der Parameter mit der Qualität der Ausgabe voraussetzen.

Eine zweite Unterscheidung nach Verwendungszweck teilt den Bereich des Data Mining in die Teilbereiche der Klassifikation, der Clusteranalyse und der Suche nach Assoziationsregeln ein. Klassifikation ist dabei, wie bereits angedeutet, gedacht als die Vorhersage der Eigenschaften von Rohdaten aus bereits eingeordneten Daten, die Clusteranalyse als die Einteilung von Rohdaten in Klassen ähnlicher Daten und die Suche nach Assoziationsregeln als das Auffinden von Zusammenhängen zwischen häufig vorkommenden Daten.

2.2.4. Bewertung der Unterscheidungsmöglichkeiten

Die Betrachtung der oben genannten, bunt gemischten Ansätze der Informatik zur Strukturierung des maschinellen Lernens ist hilfreich, sowohl um die Grundbegriffe des maschinellen Lernens kennen zu lernen als auch um die verschiedenen und teilweise widerstrebenden Absichten der Informatik aufzuzeigen und damit ein interdisziplinäres Verständnis der Problematik zu entwickeln. Nichts desto trotz sind die dargestellten Perspektiven der Informatik zur Unterscheidung von Algorithmen aus technikphilosophischer Sicht mangels Systematik nicht befriedigend. Auch die Kombination der Kriterien ist auf dem Weg zu einem systematischeren Überblick über das maschinelle Lernen nicht hilfreich. Die Unterscheidungen nach Rückmeldungen und nach Suchstrategien sind zwar weitgehend

unabhängig voneinander, ein Kreuzvergleich der beiden Unterscheidungsarten ergibt aber keine neuen Erkenntnisse. Zwar findet sich der typischere Einsatz einer parametrischen Suche im überwachten Lernen, da dort meist generell mehr Vorwissen besteht, aber auch im unüberwachten Lernen ist eine parametrische Suche denkbar. In die andere Richtung gedacht, kann bestärkendes Lernen sowohl bei einem geordneten Suchraum als auch bei einer stochastischen Suche eingesetzt werden. Gerade die Unterscheidung nach der Suchstrategie ist darüber hinaus kategorial nicht ganz einheitlich, da etwa die nichtparametrische Suche ein Bereich ist, der je nach Lesart des Begriffes sehr viel mehr enthält als nur stochastische Suchen.

Prinzipiell überrascht eine solche Gemengelage nicht, da die Definition von maschinellem Lernen sehr allgemein gehalten ist, sich also sehr viele Ansätze und Entwicklungsziele unter der Überschrift vereinen lassen und die Informatik als Disziplin nur ein eingeschränktes Interesse daran hat, den Bereich als Ganzen systematisch diskutieren zu können. Die genannten Unterscheidungen sind von Anwendungsgebieten und Rahmenbedingungen des Einsatzes von MLA motiviert und sind von den Spezifika des maschinellen Lernens erst einmal unabhängig. Zum Teil werden die Algorithmen nur noch über ihre Ergebnisse klassifiziert beziehungsweise typisiert und damit benennbar gemacht. Das ist insbesondere im maschinellen Lernen problematisch, da hier das Lernergebnis durchaus sinnvoll als Black Box betrachtet werden kann und häufig keinen Rückschluss mehr auf den Autoadaptionsprozess zulässt. Andererseits hat sich überraschenderweise in der Informatik neben oder vor der genannten Unterscheidung des Gesamtgebietes in Teilbereiche eine andere Art durchgesetzt, maschinell lernende Algorithmen zu verorten. Einzelne Algorithmen werden nach der zugrunde liegenden Idee zu Klassen zusammengefasst und das allgemeine Funktionsprinzip hinter dem Algorithmus spielt dabei eine größere Rolle als fachliche Details. Diese Klassen sind die der **Lernstrategien**.

2.3. Klassifizierung nach Lernstrategien

Die im Folgenden vorgenommene Klassifizierung von Gruppen lernender Algorithmen als Lernstrategien basiert auf dem Ansatz, maschinelles Lernen nicht vom Lernergebnis aus zu denken. Die Algorithmen sollen gerade nicht als Black Box betrachtet werden, die nach Ablauf eines opaken Autoadaptionsprozesses in der Lage sind Rohdaten Ausgabestrukturen zuzuordnen. Gleichzeitig soll die Klassifizierung nicht zu sehr auf die Details der konkreten Implementierung von MLA eingehen, denn formal entstehen im maschinellen Lernen ständig neue Algorithmen. Im Rahmen des Autoadaptionsprozesses verändert schon die Aufnahme der Eingabe- oder Sensordaten formal die Struktur des zugrunde liegenden Algorithmus. Die Klassifizierung wird stattdessen vorgenommen, indem Algorithmen der gleichen Lernstrategie zugeordnet werden, wenn die Selbstorganisationsprinzipien, die hinter den jeweiligen Autoadaptionsprozessen stehen, sich ähneln. Diese vom Algorithmus gedachte und über die Strategie argumentierende Klassifizierung wird auch innerhalb der Informatik in ähnlicher Weise vorgenommen. Die Lernstrategien sind meist gut voneinander abzugrenzen, allerdings gibt es Ausnahmen, beziehungsweise Grenzfälle, die im Einzelnen diskutiert werden, soweit sie aus interdisziplinärer Perspektive einen Mehrwert bieten. Insgesamt kann die Methodik per Konstruktion alle denkbaren Algorithmen abdecken und kann insbesondere auch mit Mischformen und neu entstehenden Strategien ohne größere Schwierigkeiten umgehen.

Die im maschinellen Lernen erzeugten Ausgabestrukturen sollen zwar nicht als Basis für die Klassifikation genutzt werden, spielen aber dennoch eine zentrale Rolle. Zwar können die erzeugten Ausgabestrukturen formal häufig als **Recommender-Systeme** und damit als Entscheidungsalgorithmen beschrieben werden, aber eine solche Interpretation erfolgt nicht zwangsläufig und ist in den meisten Fällen nicht hilfreich. Die Ausgabestrukturen von MLA werden im Weiteren stattdessen als Strukturvorschläge betrachtet und bezeichnet. Unabhängig davon, inwiefern ein MLA einen Algorithmus erstellt oder nicht, erstellt das MLA formal nie eine wiederum autoadaptive Struktur, da ein Strukturvorschlag formal erst dann vorliegt, wenn der Autoadaptionsprozess beendet oder eingefroren wurde.

Zusammengefasst denkt die Klassifizierung von Lernstrategien das maschinelle Lernen vom Algorithmus aus und interessiert sich in erster Linie für den Autoadaptionsprozess und nur nachrangig für den resultierenden Strukturvorschlag.

2.3.1. Überblick der Lernstrategien

Die Darstellung der Klassifizierung maschinell lernender Algorithmen nach Lernstrategien stellt das zentrale Element des ersten Hauptteils dar. Die Reihenfolge, in der die Lernstrategien dargestellt werden, begründet sich dabei wie folgt.

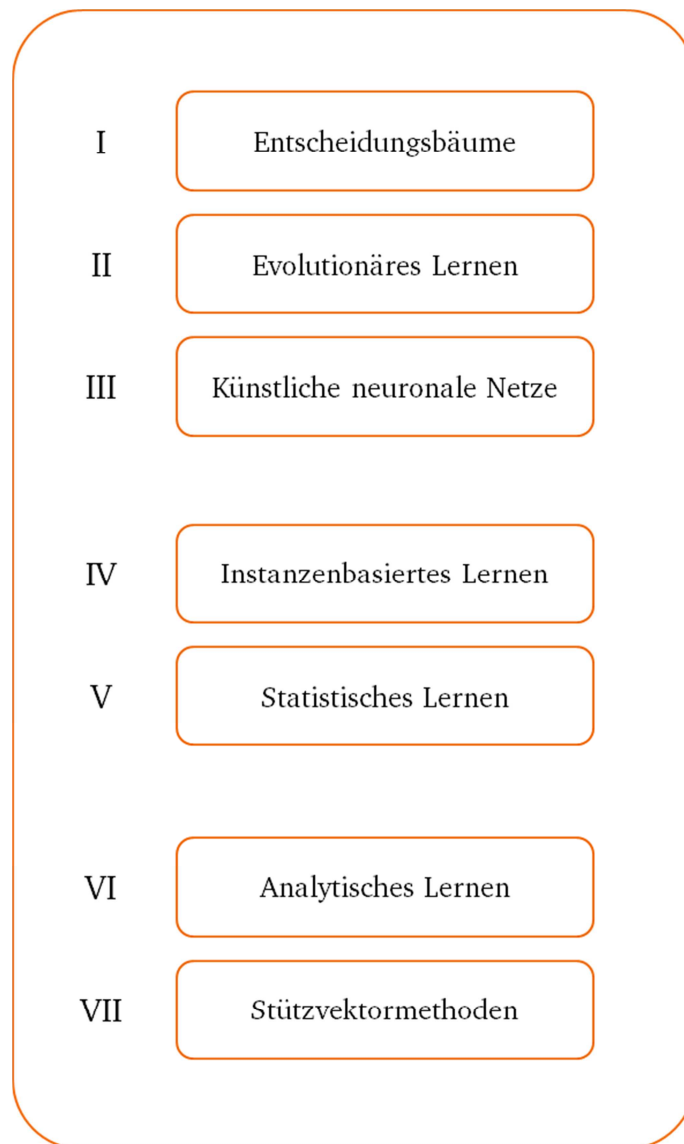


Abbildung 5: Auflösung des maschinellen Lernens nach Lernstrategien

Die erste dargestellte Klasse von lernenden Algorithmen sind die **Entscheidungsbäume**. Das Prinzip der hier verorteten lernenden Algorithmen ist einfach zu verstehen und den meisten Menschen aus anderen Kontexten als dem maschinellen Lernen bereits bekannt. Die Entscheidungsbäume eignen sich entsprechend gut als Einstiegsstrategie. Die zweite dargestellte Klasse ist das **evolutionäre Lernen**. Das evolutionäre Lernen ist eine Zusammenfassung von drei stark verwandten Lernstrategien, deren Diskussion zwar nicht unproblematisch ist, die aber alle drei gut beschreibbar sind, da nur wenige formale Formulierungen und Erklärungen notwendig sind. Die dritte dargestellte Klasse sind die **künstlichen neuronalen Netze**. Die hier betrachteten autoadaptiven Algorithmen lassen sich aus interdisziplinärer Perspektive ähnlich gut beschreiben wie evolutionäres Lernen – allerdings sind zur korrekten Darstellung künstlicher neuronaler Netze deutlich mehr formale Details notwendig. Die genannten drei Klassen erfordern zwar mitunter den Umgang mit formalen Konzepten der Informatik, allerdings wird der Zugang zu diesen Konzepten durch hilfreiche Intuitionen zu Bezeichnungen wie „evolutionäres Lernen“ erleichtert.

Die vierte und fünfte Klasse können zwar nicht auf eine entsprechende Intuition verweisen, nutzen aber eingängige und zum Teil allgemein bekannte mathematische Konzepte, die im

Rahmen der Diskussion der Unterschiede zwischen Lernstrategien bereits sehr kompakt eingeführt wurden. Die vierte Klasse bildet das **instanzenbasierte Lernen**. Das instanzenbasierte Lernen basiert stark auf der bereits kurz beleuchteten Clusteranalyse. Die fünfte Klasse des **statistischen Lernens** ist relativ eng mit dem instanzenbasierten Lernen verbunden und vereint wie schon das evolutionäre Lernen drei Lernstrategien, die stark verwandt sind.

Die sechste und siebte Klasse werden im Folgenden nur kurz skizziert. Einerseits sollte zu diesem Zeitpunkt schon ein recht gutes Verständnis für das maschinelle Lernen als Gebiet entstanden sein und andererseits spielen sie aus interdisziplinärer Sicht zunächst eine nachgeordnete Rolle²¹. Die Klasse des **analytischen Lernens** basiert wesentlich auf der Idee, direkt logische Aussagen zu manipulieren und einen geordneten Suchraum solcher Aussagen zu betrachten. Zwar bildet dieses Vorgehen einen interessanten Ansatz für die Realisierung maschinellen Lernens, die Lernstrategie spielt allerdings in der Praxis keine große Rolle. Die Klasse der **Stützvektormethoden** schließlich stellt ein Beispiel für den oben genannten Fall dar, dass neu entstandene Algorithmen, die auf bereits bekannten Lernstrategien basieren, eine neue Lernstrategie entstehen lassen. Im Fall der Stützvektormethoden waren die Algorithmen mathematisch nicht neu, aber die systematische Umsetzung im maschinellen Lernen war es. Die Stützvektormethoden werden darüber hinaus ein besonders gutes Beispiel für Algorithmen darstellen, die im zweiten Hauptteil nicht von weiterem Interesse sind und können somit zur Abgrenzung genutzt werden.

Im praktischen Einsatz und bei der Erstellung von MLA wird zwar eine Vielzahl von Hybriden der oben beschriebenen Strategien eingesetzt, die These ist jedoch, dass eine Kombination verschiedener Lernansätze und Weiterentwicklungen in den allermeisten Fällen keine prinzipiell neuartigen Verhaltensweisen von und Interaktionsformen mit lernenden Maschinen entstehen lässt, die nicht nur eine Kombination beziehungsweise Überlagerung der im Weiteren beschriebenen Verhaltensweisen darstellt. Tritt doch der Fall auf, dass unerwartete Effekte bei der Konstruktion von Mischformen beobachtet werden, würde dies zu der Entstehung einer neuen Lernstrategie führen, die auf der Erzeugung beziehungsweise Nutzung des entsprechenden Phänomen beruhen würde. Unabhängig von neu entstehenden Strategien, decken die genannten Lernstrategien das aktuelle vorliegende maschinelle Lernen zu sehr großen Teilen ab. Weitere Lernstrategien werden zwar ständig entwickelt und verworfen, aber das Ziel des ersten Hauptteils liegt darin, ein interdisziplinäres Verständnis der etablierten Technik des maschinellen Lernens zu vermitteln. Hierfür ist ein gutes Verständnis der genannten Lernstrategien mehr als ausreichend.

Die nachfolgenden Darstellungen sollen auch dazu dienen, bei der Arbeit an technikphilosophischen Fragen schnell auf die hier geleistete Grundlagenarbeit zugreifen zu können. Aus diesem Grund und zur besseren Lesbarkeit werden mitunter bereits eingeführte Begriffe noch einmal kurz erläutert.

2.3.2. Lernen von Entscheidungsbäumen

Funktionsbeschreibung

Ein **Entscheidungsbaum** ist eine Struktur, die **Eingabedaten** aufnimmt, die durch eine Menge von **Attributen** beziehungsweise Attributwerten vollständig beschreibbar sein müssen.

²¹ Insbesondere spielen beide in der Diskussion des zweiten Hauptteils keine große Rolle.

Ein Entscheidungsbaum prüft für jede Eingabe eine Anzahl von Kriterien und führt anhand der Prüfungsergebnisse eine **Klassifizierung** der übergebenen Struktur durch. Diese Klassifizierung wird ausgegeben. Die gemeinsame Visualisierung aller Prüfungskriterien wird Entscheidungsbaum genannt. Entsprechend handelt es sich bei einem Entscheidungsbaum um einen **Klassifikator**, bei dem die möglichen Klassen typischerweise vorgegeben sind und die Aufgabe nur darin besteht Prognosen abzugeben, welcher Klasse ein gegebenes Eingabedatum zugeordnet werden soll.

Die Betrachtung von Entscheidungsbäumen als Klasse von Algorithmen und damit als Lernstrategie bezieht sich auf die Konzeption eines Autoadaptionsprozesses, der mit dem Ziel gestartet wird, die Fähigkeit zur Klassifikation von Eingabedaten zu ermöglichen. Wenn der Entscheidungsbaum als Modell vorliegt, wird der Autoadaptionsprozess als abgeschlossen betrachtet. Wenn geplant ist, den Entscheidungsbaum möglicherweise zukünftig noch einmal zu adaptieren, wird der Lernvorgang als **eingefroren** bezeichnet. Nicht der Einsatz eines Entscheidungsbaumes zur Klassifizierung von Eingabedaten, sondern die Modellbildung, die einem Entscheidungsbaum vorangeht, ist der Vorgang, der als maschinelles Lernen betrachtet wird. Die fertigen Entscheidungsbäume stellen dementsprechend das Ergebnis des Einsatzes eines auf einer gewissen Lernstrategie basierenden lernenden Algorithmus beziehungsweise MLA dar.

Beispiel für einen Entscheidungsbaum

Ein Entscheidungsbaum in seiner finalen Form kann ohne größere Schwierigkeiten in einer für den Nutzer direkt lesbaren Form dargestellt werden. Derartige Strukturvorschläge werden als **symbolische Systeme** bezeichnet. Ein **subsymbolisches System** hingegen ist ein System oder Modell, dessen Funktion von einem menschlichen Betrachter nicht ohne größere Schwierigkeiten erkannt werden kann²². Die folgende Illustration zeigt ein Beispiel für einen Entscheidungsbaum, der die Frage beantwortet, wo zu Mittag gegessen werden kann.

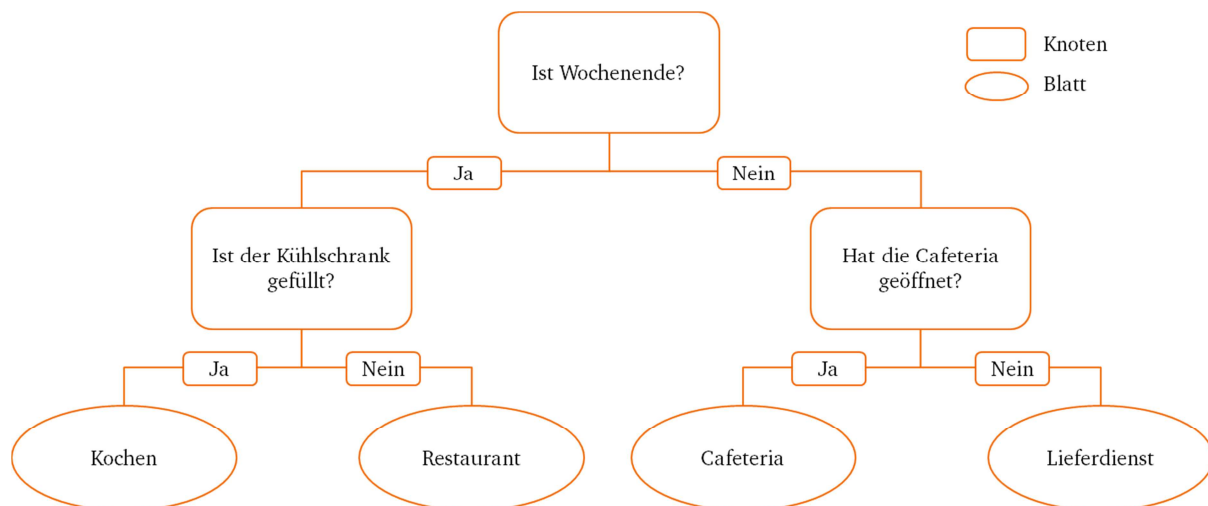


Abbildung 6: Entscheidungsbaum zur Wahl des Mittagessens

Die betrachteten Attribute der Eingabedaten sind hier der Wochentag, die Nutzbarkeit der Cafeteria und die Befüllung des Kühlschranks. Die Eingabedaten können noch weit mehr

²² Die Unterteilung in symbolische und subsymbolische Systeme ist eine weitere verbreitete Möglichkeit lernende Algorithmen zu unterscheiden. Jedes symbolische System kann jedoch durch eine Kodierung in ein subsymbolisches System umgeformt werden.

Attribute aufweisen, die Suche der relevantesten Attribute ist fast immer Teil der Problemstellung. Die auf die Attribute bezogenen Prüfungskriterien, die der Entscheidungsbaum untersucht, sind auf **Knoten** festgehalten und die möglichen Prüfungsergebnisse entsprechen **Kanten**, das heißt Verbindungen von Knoten zu anderen Knoten oder zu **Blättern**. Blätter stellen „Endpunkte“ des Entscheidungsbaums dar, an denen eine Entscheidung getroffen und damit eine Klassifizierung vorgenommen wird. Blätter weisen den auf ihnen noch betrachteten Eingabedaten Klassen zu. Diese Zuweisungen können **statisch** sein, indem wie im obigen Beispiel eine Klasse festgelegt wird, oder **dynamisch** mittels Funktionen durchgeführt werden. Das Blatt links unten im Beispiel könnte etwa die zu kochende Menge abhängig von den bereits verspeisten Kilokalorien vorgeben. Diese Verwendung von **Funktionen** auf Blättern ließe sich im Prinzip auch durch weitere Knoten und Blätter ersetzen, allerdings würde die Übersichtlichkeit des Baumes sehr leiden, wenn tausende Blätter ergänzt würden, deren korrespondierende Klassen sich nur um eine Kilokalorie unterscheiden. Mit diesem Trick können auch Klassifizierungsprobleme bearbeitet werden, die eine Unterscheidung zwischen unendlich vielen Klassen erfordern.

Praktische Anwendungsbeispiele umfassten in der Vergangenheit beispielsweise die Bewertung des Risikos bei der Auswahl von Anwärtern für Kredite oder die Erstellung von medizinischen Diagnosen (Mitchell 1997).

Konstruktion eines Entscheidungsbaumes

Zu einer **Problemstellung** in Form einer vorzunehmenden Klassifizierung sei eine Anzahl von **Trainingsdaten** gegeben, deren Klassenzugehörigkeit bereits bekannt ist – das heißt, eine Anzahl von **Instanzen** im eigentlichen Sinn. Die Betrachtung und Prüfung eines oder mehrerer Attribute einer Instanz wird als **Test** bezeichnet. Jeder Test entspricht einer Klassifizierung, da die getesteten Daten auf mindestens zwei nachfolgende Knoten aufteilt werden. Ein Entscheidungsbaum kann diese Aufspaltung so intensiv betreiben und so viele Tests vorschreiben, dass jedes Blatt nur genau eine Trainingsinstanz beschreibt – das wird notwendig, wenn jede Instanz jeweils der einzige **Repräsentant** einer Klasse ist.

Die Entscheidungskriterien der Tests können **stochastisch** sein, das heißt ein Prüfungsergebnis wird nur mit einer gewissen Wahrscheinlichkeit einem Datum zugeordnet²³. Tests werden, wie im Beispiel zu sehen, als Knoten visualisiert und der Knoten des ersten durchzuführenden Tests wird als **Wurzelknoten** bezeichnet.

Die Erstellung eines Entscheidungsbaumes folgt nun – beginnend mit der Betrachtung des Wurzelknotens – den folgenden Schritten.

- A** Falls die am betrachteten Knoten noch verbliebenen Trainingsdaten sich vollständig aus Instanzen einer einzelnen Klasse zusammensetzen, wird der betrachtete Knoten zu einem **Blatt** und die Schritte **B** und **C** werden ausgelassen.
- B** Für alle Tests, die für die Instanzenmenge am betrachteten Knoten durchgeführt werden könnten, wird geprüft, wie gut der Test die Instanzen trennt. Die Kenngröße wird allgemein als **Qualität** bezeichnet, kann jedoch unterschiedlich konstruiert sein. Der Test mit der höchsten Qualität wird dem Knoten zugeordnet.

²³ Hierdurch entscheidet der Entscheidungsbaum bei wiederholter Eingabe des gleichen Datums unterschiedlich.

Der Knoten teilt damit die betrachteten Eingabedaten entsprechend der Testergebnisse in disjunkte Teilmengen ein.

- C Für jede dieser disjunkten Teilmengen wird ein neuer Knoten erzeugt, der nur die Instanzen der jeweiligen Gruppe testet. Diese neuen Knoten werden als **Kinderknoten** bezeichnet und über eine Kante mit dem erzeugenden Knoten verbunden.
- D Der Prozess beginnt für einen noch nicht betrachteten Knoten erneut bei Schritt A. Wenn keine noch nicht betrachteten Knoten verblieben sind, ist die Erstellung des Baums abgeschlossen.

Sind alle Eingabedaten Instanzen einer einzigen Klasse, so besteht der Entscheidungsbaum nur aus einem Blatt und keinem Knoten. Wenn die Eingabedaten Instanzen aus zwei Klassen enthalten, kann der Entscheidungsbaum mit einem Knoten und zwei Blättern auskommen. Die Qualität des Tests wäre in diesem Fall maximal. Der Zusammenhang des Konzepts von Qualität bei der Erstellung von Entscheidungsbäumen wird später noch betrachtet. In der Praxis ist die Identifikation eines einzigen Tests, der alle Eingabedaten eindeutig in Klassen einordnet, meist nicht möglich und es müssen eine Reihe von Tests durchgeführt werden um zwei Klassen von Instanzen voneinander zu **trennen**. Ein Entscheidungsbaum darf nur endlich viele Knoten enthalten, dies muss unter anderem bei der Wahl des Qualitätskonzeptes für Schritt B sichergestellt werden, etwa indem jedes Attribut der Eingabedaten nur genau einmal im Rahmen eines Tests überprüft werden darf. Üblicherweise werden Entscheidungsbäume daher, trotz der Möglichkeit Blätter dynamischer Klassenzuordnung zu verwenden, nur zur Klassifizierung genutzt, wenn eine endliche Anzahl von Trainingsdaten vorliegt und nach endlich vielen Klassen getrennt werden soll.

Der Qualitätsbegriff bei Entscheidungsbäumen

Das Wesentliche an Entscheidungsbäumen ist aus Sicht der Lernstrategien des maschinellen Lernens nicht der Entscheidungsbaum selbst, sondern der Autoadaptionsprozess, der diesen Entscheidungsbaum entstehen lässt. Das wichtigste Konzept bei der Erstellung dieser Entscheidungsbäume wiederum ist das Konzept der **Qualität** eines Tests. Die Darstellung des Qualitätsbegriffes basiert unter anderem auf dem Verständnis von **Information** innerhalb des maschinellen Lernens und erfordert eine gewisse Vorarbeit. Innerhalb des maschinellen Lernens ist fast immer die syntaktische Ebene gemeint, wenn von Informationen die Rede ist.

Der Informationsgehalt einer Nachricht entspricht der Anzahl der Ja-/Nein-Fragen, die man bei einer idealen Fragestrategie braucht, um sie zu rekonstruieren.

(Wikipedia Contributors 2012, Information)

Der Begriff der Information ist im Kontext des maschinellen Lernens zwar klar bestimmt, dient jedoch in erster Linie der Bestimmung des wesentlicheren und innerhalb des maschinellen Lernens ebenso klar bestimmten Begriffs der **Entropie**. Die Entropie misst die Durchmischung von Trainingsinstanzen, die positive und negative Beispiele einer Klassifizierung darstellen. Wenn alle Trainingsdaten der gleichen Klasse angehören, ist diese Durchmischung beziehungsweise die Entropie minimal. Wenn hingegen genau gleich viele Trainingsdaten den jeweils vorliegenden Klassen angehören, ist die Entropie maximal. Die

Reduktion der Entropie in Datenbeständen ist häufig die Motivation zum Einsatz von MLA und das Konzept von Informationen spielt nur insofern eine Rolle, als es dem Informationsgewinn und damit der Messung der Entropie zugrunde liegt. Die Qualität eines Entscheidungsbaumes misst die Verbesserung der Entropie, die mit Hilfe eines speziellen Tests erzielt wird. Diese Verbesserung wird als **Informationsgewinn** bezeichnet und gibt an, wie stark die durch den Test neu entstehenden Teilmengen mit Instanzen aus verschiedenen Klassen durchmischt sind, relativ zur Durchmischung der vor dem Test vorliegenden Menge von Instanzen. Angestrebt wird die Entstehung von Teilmengen, deren Elemente jeweils genau einer Klasse angehören, das heißt die Entstehung von Blättern. Blätter entstehen, wenn eine Durchmischung von null bei einer durch den Test neu entstehenden Teilmenge getesteter Instanzen festgestellt wird. Das Konzept des Informationsgewinns bringt die Gefahr mit sich, dass einzelne Knoten die betrachteten Daten **fragmentieren**, das heißt im Extremfall sämtlichen betrachteten Instanzen isolierte, individuelle Blätter zuordnen. Eine Weiterentwicklung des Qualitätskonzeptes zum Umgang mit diesem Problem ist das Konzept des **Gewinnverhältnisses**. Hier wird der Grad, in dem die Instanzenmenge durch den Test fragmentiert wird, in das Qualitätskonzept integriert. Zusammengefasst stellt die Entropie die Kennzahl zur Messung der Fragmentierung einer Datenmenge dar und die Sicherstellung einer minimalen Entropie in der **Partitionierung** einer Instanzenmenge wird als **Informationstrennung** bezeichnet. Eine andere Weiterentwicklung des Informationsgewinns berücksichtigt die Tatsache, dass gewisse Tests nur unter sehr hohen Kosten durchzuführen sind oder längere Zeit benötigen. Zur Berücksichtigung dieser Umstände kann bei der Bestimmung der Qualität eines Tests eine Strafe für teure Tests aufgenommen werden. Die an dieser Stelle eingeführten technischen Ausdrücke zur Arbeit mit dem Begriff der Information werden im Weiteren keine explizite Verwendung finden. Die Einführung dieser Ausdrücke ist dennoch hilfreich, da sie die Diskussion von Informationstechnik in Abschnitt 3.3 vorbereitet.

Vorteile und Nachteile von Entscheidungsbäumen

Im Folgenden werden die wesentlichsten Stärken und Schwächen von MLA beschrieben, die als Ausgabestruktur Entscheidungsbaume erstellen. Zwar ist der entstehende Entscheidungsbaum als Lernergebnis und damit als Strukturvorschlag des Autoadaptionsprozesses für das Verständnis ebenjenes Prozesses formal nicht von höchster Bedeutung, aber die Grundidee dieser Lernstrategie basiert darauf, dass Entscheidungsbäume manipuliert werden sollen. Entsprechend wird im Weiteren nicht explizit zwischen Vor- und Nachteilen des Autoadaptionsprozesses und des resultierenden Strukturvorschlages unterschieden. Tatsächlich gehen beide Bereiche bei der Entwicklung der konkreten Algorithmen ineinander über, da der Lernprozess, wie oben beschrieben, schrittweise Knoten und Blätter hinzufügt, wodurch er den entstehenden Entscheidungsbaum vergrößert. Das MLA muss somit in jedem Schritt mit der Struktur arbeiten, die auch das Endergebnis darstellt. Das heißt, eine schnelle Reaktionsgeschwindigkeit von Entscheidungsbäumen beschleunigt auch den Lernprozess. Dieser Zusammenhang ist für andere Lernstrategien, insbesondere die künstlichen neuronalen Netze, ebenfalls von großer Bedeutung. Die Rede von einem Autoadaptionsprozess findet hier ihre wesentlichste Motivation. Die iterative Adaption des Entscheidungsbaums ist die Adaption der Struktur, die aus Reizen gelernt hat. Das Verhältnis des Autoadaptionsprozesses und der veränderten Struktur wird im Rahmen der Diskussion nichttrivialer Maschinen im zweiten Hauptteil weiter ausgeführt.

Gut geeignet für die Konstruktion eines Entscheidungsbaumes sind vor allem Problemstellungen, bei denen eine für die Nutzer schnell verständliche Darstellung der

getrennten Klassen bevorzugt wird. Auch erfordert die Klassifizierung eines Eingabedatums nur eine vergleichsweise kurze **Laufzeit**, also Bearbeitungsdauer der Klassifizierung durch das MLA. Die Nutzung von Entscheidungsbäumen in der Praxis ist entsprechend unproblematisch.

Die Entwicklung von Entscheidungsbäumen ist **robust** gegen **Rauschen**. Das bedeutet, fehlerhafte Messwerte und damit Trainingsdaten, die nur annähernd korrekte Aussagen machen, behindern den Autoadaptionprozess nur geringfügig. Entscheidungsbäume können zudem auch dann gelernt werden, wenn Trainingsdaten unvollständig sind, das heißt wenn Werte von Attributen fehlen, etwa weil Aufzeichnungen fehlen oder zu teuer waren²⁴.

Ein Hauptnachteil bei der Entwicklung von Entscheidungsbäumen liegt darin, dass der Lösungsraum **unvollständig durchsucht** wird, wodurch Fehlklassifikationen entstehen können. Die Bezeichnung als unvollständige Suche bezieht sich darauf, dass durch die einmalige Entscheidung für einen speziellen Test für jeden Knoten und die darauf aufbauende Erweiterung des Modells all diejenigen Entscheidungsbäume nicht betrachtet werden, bei denen der jeweilige Test ein anderer wäre. Ein weiterer wesentlicher Nachteil liegt in der Gefahr einer **Überanpassung** an die Trainingsinstanzen. Von Überanpassung wird gesprochen, wenn ein MLA die Trainingsdaten zu genau berücksichtigt und beispielsweise Attribute, die keine Rolle spielen, in Tests miteinschließt. Wenn beispielsweise Sehenswürdigkeiten auf Basis von Fotos identifiziert werden sollen und auf allen Trainingsinstanzen, die den Eiffelturm zeigen, eine Wolke am Himmel ist, könnte der Entscheidungsbaum diese Eigenschaften als Attribut abprüfen, bevor er eine Abbildung als den Eiffelturm erkennt.

Abgesehen von den beispielhaft aufgezählten Vor- und Nachteilen weisen Entscheidungsbäume einen prinzipiellen **Bias** – eine konstruktionsbedingte systematische Verzerrung – auf. Dieser Bias wird innerhalb der Informatik vielfach als induktive Verzerrung bezeichnet, wenngleich der Begriff fast ausschließlich auf Englisch verwendet wird. Diese Bezeichnung ist jedoch irreführend, gemeint ist eine **abduktive Verzerrung** bei der Qualitätsbewertung²⁵. Die abduktive Verzerrung besteht in einer Bevorzugung von kleinen gegenüber großen Entscheidungsbäumen und damit kompakten gegenüber umfangreichen Strukturvorschlägen. Insbesondere werden Entscheidungsbäume, die einen hohen Informationsgewinn nahe dem Wurzelknoten aufweisen, bevorzugt. Dies führt systematisch zu weniger Überanpassung, allerdings gegebenenfalls auch zu theoretisch auf Basis der Trainingsinstanzen vermeidbaren Fehlklassifizierungen – Details zu dem Problem der unvollständigen Suche und der abduktiven Verzerrung werden in der Diskussion des Stutzens erläutert.

Stutzen als entscheidungsbaumspezifische Maßnahme gegen Überanpassung

Entscheidungsbäume können im finalen Zustand oder in Zwischenzuständen während des Autoadaptionprozesses **gestutzt** werden, um eine Überanpassung zu vermeiden. Eine Stutzung von Entscheidungsbäumen wird häufig durchgeführt, indem die Kaskade von einem Knoten über all dessen Kinderknoten bis hin zu den Blättern – wenn man so will ein **Ast** – durch ein Blatt ersetzt wird. Das neu erstellte Blatt kategorisiert Eingabedaten entsprechend

²⁴ Formal gesprochen sind Entscheidungsbäume in der Lage sehr verschiedene Strukturvorschläge zu lernen. (Präzise: Sie können unter anderem alle Funktionen modellieren, die auf endlichen diskreten Mengen operieren, sind vollständig ausdrucksstark in der Klasse der aussagenlogischen Sprachen (Russell et al. 2007) und können zu Kausalsätzen umgeschrieben werden.)

²⁵ Details zur Verwendung der Begriffe Induktion und Abduktion in der Informatik finden sich bei Kaminski und Harrach (Kaminski et Harrach 2010).

der im gestutzten Ast vorrangigen Klassifikation. Häufig wird dabei überprüft und sichergestellt, dass der entstehende, verkleinerte Entscheidungsbaum die Trainingsdaten mindestens genauso präzise klassifiziert wie der ungestutzte Entscheidungsbaum. Die Fähigkeit eines MLA zur präzisen Klassifizierung wird als Betrachtung der **Performanz** bezeichnet, eine Stutzung soll entsprechend keinen Performanzverlust auf den Trainingsdaten mit sich bringen²⁶.

Eine alternative Maßnahme, um gegen Überanpassung oder zu große Komplexität des Entscheidungsbaumes vorzugehen, ist die **Regelstutzung**. Hier wird genutzt, dass sich jedes Blatt und damit jede Klassifizierung als ein Kausalsatz darstellen lässt. Im obigen Beispiel wäre der Kausalsatz für das rechte untere Blatt „wenn weder Samstag noch Sonntag ist und die Cafeteria geschlossen hat, dann wird beim Lieferdienst bestellt“. Eine Regelstutzung formuliert für jedes Blatt den Weg vom Wurzelknoten als einen solchen Kausalsatz und entfernt anschließend eine der Voraussetzungen aus dem Kausalsatz. Eine Möglichkeit im Beispiel des rechten unteren Blattes wäre „wenn die Cafeteria geschlossen hat, dann wird beim Lieferdienst bestellt“. Hier ist gut zu erkennen, dass in diesem Fall eine Regelstutzung die Performanz des Baumes stark beschädigt. Dieser Effekt ist auch nicht überraschend, da der Baum schon sehr kurz war und kein Ast redundant erscheint.

Generell ist bei der Reduktion der Analysetiefe von MLA zum Kampf gegen Überanpassung – insbesondere beim Stutzen von Entscheidungsbäumen – eine Dynamik in der Performanz zu beobachten. Auch wenn sich durch eine Stutzung die Performanz insgesamt nicht verschlechtert, kann dadurch dennoch eine erhebliche Anzahl von neuen Fehlklassifikationen erzeugt werden, solange die Anzahl der vermiedenen Fehlklassifikationen noch größer ist. Eine vom Nutzer hingenommene Erzeugung von zusätzlichen Fehlklassifikationen kann beispielsweise auftreten, wenn die Trainingsinstanzen widersprüchlich sind und ein teilweise irrtümlich entstandener Ast gestutzt wird. In diesem Fall wird statt des Astes ein Blatt mit der mehrheitlich richtigen Klasse eingefügt. Somit werden einige der Eingabedaten, die bisher richtig klassifiziert wurden, in der gestutzten Variante des Gesamtbaumes falsch zugeordnet. Eine Idee dahinter ist, dass häufig unsystematische Fehler oder Rauschen in den Trainingsdaten zu filigranen Verästelungen führen und sich die Nutzer nicht für eine Nachbildung des zufälligen Rauschens interessieren.

Der Vorgang des Stutzens und die Gefahr der Überanpassung im Allgemeinen sollen nachfolgend an einem Beispiel veranschaulicht werden. Die Beispielaufgabe sei eine Klassifikation von Bankkunden als kreditwürdig oder als nicht kreditwürdig. Es sei angenommen, dass in den Trainingsdaten genau zehn Kunden enthalten sind, die parallel zwei Kredite abbezahlen. Es sei weiter angenommen, dass neun von diesen zehn Kunden mit beiden Krediten im Verzug sind und dass der zehnte Kunde bei genau einem der Kredite zahlungsfähig und außerdem schwerhörig ist.

²⁶ In der Informatik werden neben der Performanz noch andere Parameter eines Algorithmus optimiert. Beispiele sind der benötigte **Speicherplatz** oder die **Laufzeit**. Diese Parameter sind interdisziplinär jedoch nur von geringem Interesse und werden im Weiteren nicht explizit betrachtet.

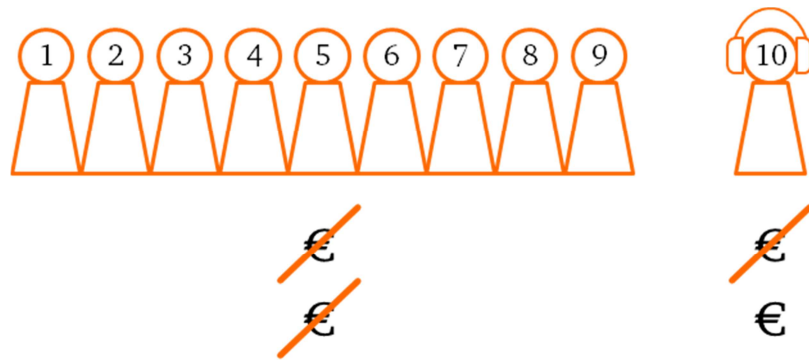


Abbildung 7: Trainingsdaten des Beispiels zur Kreditwürdigkeit

Ein Entscheidungsbaum wird daraufhin wahrscheinlich bei der Frage nach der Vergabe eines zweiten Kredites den Test „Ist der Kunde schwerhörig?“ einfügen. Dies würde auf den Trainingsinstanzen zu einer Verbesserung der Performanz führen, jedoch erscheint dieses Kriterium dennoch die Verallgemeinerbarkeit des Baumes zu gefährden. In diesem Fall würde ein entsprechendes Stutzen des Entscheidungsbaumes zu einer Verschlechterung auf den Trainingsinstanzen führen, jedoch zu einer besseren Performanz auf den späteren Eingabedaten. Natürlich kann es Kunden geben, die aus speziellen Gründen im Verzug sind und dennoch kreditwürdig wären, aber die Hörfähigkeit der Person ist diesbezüglich sehr wahrscheinlich kein geeignetes Entscheidungskriterium. Grafisch bedeutet dies, dass aus dem Entscheidungsbaum A der gestutzte Entscheidungsbaum B wird.

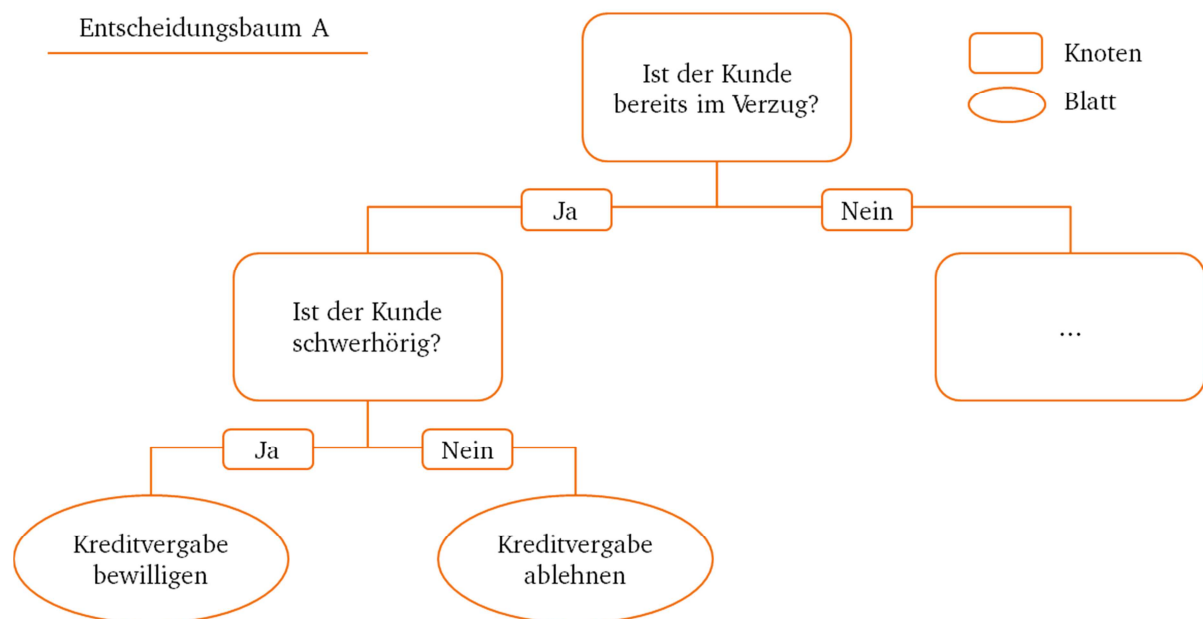


Abbildung 8: Ungestutzter Entscheidungsbaum

Entscheidungsbaum B

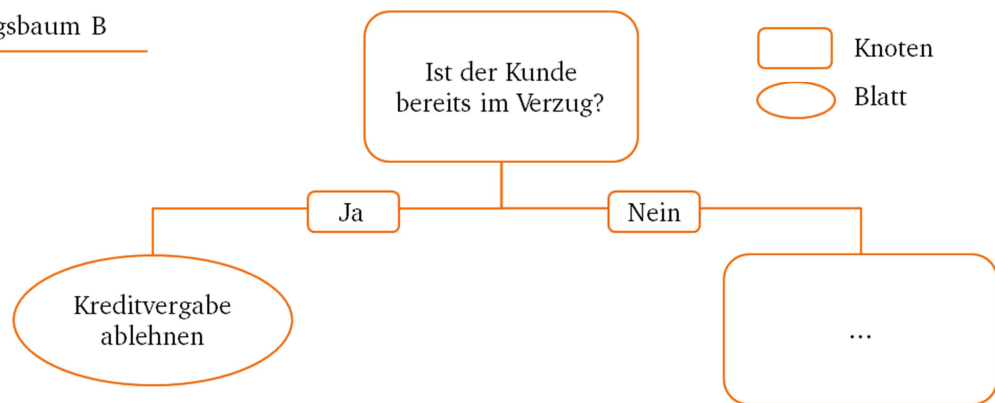


Abbildung 9: Gestutzter Entscheidungsbaum

Im Beispiel der Kreditvergabe war zu sehen, dass Entscheidungsbäume die Eigenschaft haben, dass Randphänomene, die nur wenige Instanzen betreffen, leicht falsch klassifiziert werden können. Eine Möglichkeit dies zu verbessern besteht in der Messung der Performanz auf **Testdaten** und **Validierungsdaten**. Die Trainingsinstanzen werden hier vor Beginn des Autoadaptionsprozesses in drei Teilmengen aufgeteilt²⁷. Die Teilmenge der Testdaten wird indirekt für den Lernvorgang genutzt, etwa um zu prüfen, ob der Strukturvorschlag des MLA eine Überanpassung zeigt. Die Teilmenge der Validierungsdaten wird gar nicht als Rückmeldung innerhalb des Autoadaptionsprozesses verwendet, sondern dient dazu, den Strukturvorschlag als das Endergebnis des Autoadaptionsprozesses auf seine Funktionstüchtigkeit zu überprüfen.

Ergänzende Weiterentwicklungen von Entscheidungsbäumen

Die Performanz von Entscheidungsbäumen kann erhöht werden, indem ein Gremium verschiedener Entscheidungsbäume erstellt wird, das die Aussagen der beteiligten Entscheidungsbäume bündelt und das Resultat bei der Klassifizierung von neuen Eingabedaten nutzt. In diesem Fall spricht man von **Entscheidungswäldern**.

Abschließend soll noch einmal betont werden, dass Entscheidungsbäume überraschend aussagestark sind. Sie sind, wie oben angedeutet, in der Lage, alle Kausalsätze ausdrücken und können auch **stetige** Ein- und Ausgaben verarbeiten. Als stetig wird eine Eingabe bezeichnet, bei der der konkrete Wert aus einem Kontinuum von Werten – einem **Intervall** – stammen kann und nicht auf eine endliche Auswahl beschränkt ist. Die Menge der Vornamen aller lebenden Menschen etwa ist groß, bildet aber eine diskrete Eingabemenge, während die Menge aller Zahlen zwischen null und eins eine stetige und unendlich große²⁸ Eingabemenge darstellt. Entscheidungsbäume, die stetige Ausgabewerte bewältigen können, werden als **Regressionsbäume** bezeichnet. Im Vorherigen wurde bereits angedeutet, dass Blätter mit Hilfe von Funktionen Klassifizierungen vornehmen können und dadurch stetige Ausgaben erzeugen können, ohne unendlich viele Blätter zu benötigen. Die endliche Anzahl von Blättern hat den Vorteil, dass die Erstellung von lokalen Modellen sehr intuitiv möglich ist.

²⁷ Die dritte Teilmenge bilden die Trainingsdaten selbst.

²⁸ Die Zahlen zwischen null und eins sind nicht beliebig groß, aber sie können eine beliebig große Zahl von Nachkommastellen aufweisen, und somit sind unendlich viele unterschiedliche Eingabedaten möglich.

Das heißt, die Rohdaten werden getrennt und die Hoffnung ist, dass der entsprechende Test nur solche Daten derselben Kategorie zuordnet, die auch mit demselben lokalen Modell beschrieben werden können. Der Trick, der stetige Eingaben möglich macht, liegt darin, die Eingaben als **Repräsentanten** von Intervallen zu betrachten. Ein Beispiel sei die Bestimmung der Vermehrungsrate einer bisher unbekannten Art von Ameisen. Zu diesem Zweck soll ein MLA konstruiert werden, das auf der adaptiven Erstellung eines Entscheidungsbaums beruht. Eine große Anzahl von Ameisenkolonien wird gegründet und pro Kolonie wird in regelmäßigen Abständen die verstrichene Zeit in Verbindung mit der jeweiligen Anzahl der Tiere festgehalten. Das MLA soll nach Abschluss des Autoadaptionsprozesses einen Vorschlag für die Struktur der Vermehrung machen. Formal ist die Anzahl von Ameisen endlich, allerdings wird sie schnell sehr groß, das bedeutet, die Eingabe des Strukturvorschlages würde wahrscheinlich Intervalle nutzen. Die Vermehrungsrate beschleunigt sich wahrscheinlich mit zunehmender Anzahl der Tiere bis zu einem natürlichen Maximum der Produktion von Eiern einer Königin. Das heißt, die Verwendung von lokalen Modellen für unterschiedliche Zeitintervalle scheint erfolgsversprechend. Diese lokalen Modelle sind wahrscheinlich sehr viel einfacher darzustellen als ein **globales Modell**, das alle Fälle erfasst. Der Strukturvorschlag könnte beispielsweise drei unterschiedliche Modelle für die Vermehrung erlernen, eines für die Zeit, in der die Königin noch jung ist und das Nest etabliert wird, eines für die Zeit, in der sie ungestört Eier legen kann und eines für die Zeit, in der die Königin alt ist und andere Königinnen schlüpfen.

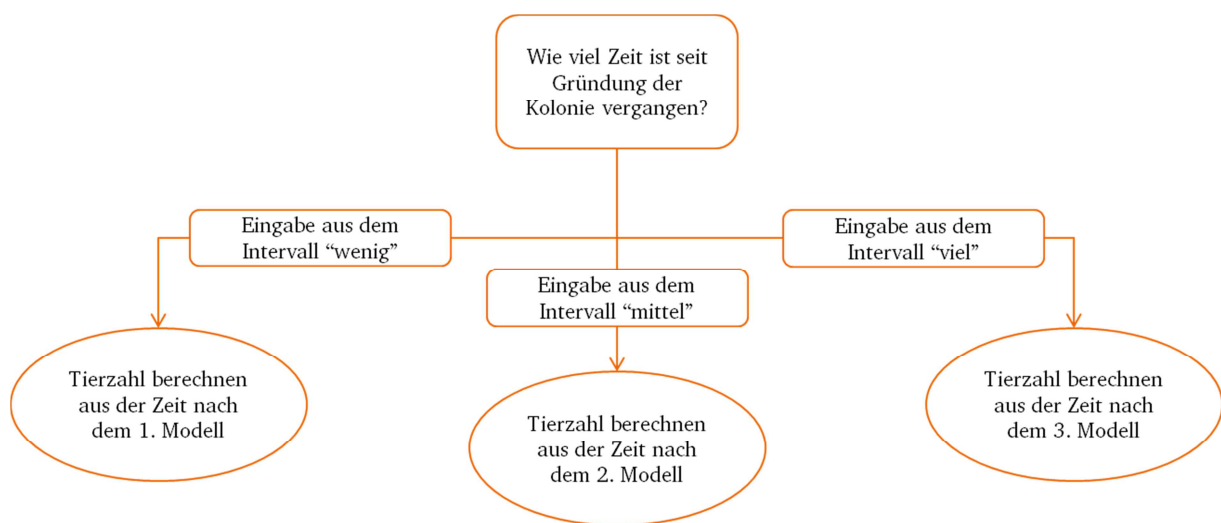


Abbildung 10: Strukturvorschlag in Form eines Regressionsbaumes

Das obige Beispiel für einen Entscheidungsbaum ist ein untypisches Beispiel für einen Vertreter dieser Klasse von Lernstrategien. Dennoch ist es illustrativ, um zu zeigen, dass sich das Konzept eines MLA, das auf Entscheidungsbäumen basiert, schrittweise in den Vertreter einer anderen Lernstrategie umwandelt, wenn die eigentliche Idee und Intuition hinter der Lernstrategie keine Rolle mehr spielt. Im Fall der Ameisen liegt ein Großteil des Aufwands in der Bestimmung der lokalen Modelle und der Wahl der passenden Intervalle. Die Erstellung des Entscheidungsbaums ist nicht ganz so komplex, kann aber durchaus innerhalb einer größeren Zahl von Modellen und Intervallen diejenigen identifizieren, die für die Tests der Knoten am geeignetsten sind. Natürlich gibt es weniger eindeutige Beispiele, bei denen ein recht komplexer Entscheidungsbaum in Kombination mit stetigen Ausgaben eingesetzt wird. In der Praxis werden sehr häufig mehrere Lernstrategien in Kombination eingesetzt, allerdings ist es auch in diesen Fällen wichtig und möglich zumindest eine Intuition zu erlangen, was

welche Lernstrategie zum entstehenden MLA beiträgt. Die Hauptstärke von Entscheidungsbäumen in diesem Zusammenhang besteht darin, dass sie eine sehr gut nutzbare Schnittstelle zwischen Nutzer und dem Autoadaptionsprozess aufweisen.

2.3.3. Evolutionäres Lernen

Die zweite Klasse von Lernstrategien, die diskutiert werden soll, ist die des **evolutionären Lernens**. Wie bereits angedeutet wurde, setzt sich die Klasse des evolutionären Lernens aus drei stark verwandten Lernstrategien zusammen, deren Diskussion zu großen Teilen vereinheitlicht geführt werden kann. Alle drei nachgeordneten Lernstrategien sind an dieser Stelle gut beschreibbar, da im Rahmen der Darstellung der Entscheidungsbäume die meisten benötigten Grundbegriffe des maschinellen Lernens schon eingeführt wurden.

Motivation der Lernstrategien des evolutionären Lernens

Die drei Lernstrategien des evolutionären Lernens modellieren ihre Autoadaption anhand der Idee eines Evolutionsprozesses als Konzept zur Anpassung von **Populationen** an Umwelteinflüsse. Wohlgedacht wird dabei keine Aussage darüber gemacht, wie Evolution tatsächlich verläuft, sondern verbreitete Ideen wie das Überleben des Stärksten werden genutzt. Die Identifizierung von informatikfernen Modellen und die mathematische Nutzbarmachung von Teilelementen in der Informatik ist ein wesentliches Konstruktionsmerkmal verschiedener Teilbereiche des maschinellen Lernens. Schon die Bezeichnung von Algorithmen als lernend war eine solche Begriffsübertragung, die allerdings das maschinelle Lernen nicht weiter beeinflusst. Zentral wird diese Vorgehensweise auch bei künstlichen neuronalen Netzen, die mit der Funktionalität von Neuronen im Gehirn ebenfalls einen natürlichen Prozess als Vorbild nehmen und diesen dann gerade *nicht* im Detail nachbilden, sondern eine begrenzte Auswahl von Prozessbausteinen als Inspiration nutzen, um daraus einen genuin der Informatik entspringenden Ansatz zu entwerfen. Das bedeutet, evolutionäres Lernen modelliert keine faktisch ablaufende evolutionären Prozesse, genau wie künstliche neuronale Netze keine Gehirne modellieren und statistische Lernverfahren keine statistischen Methoden einsetzen. Stattdessen formuliert evolutionäres Lernen mit Hilfe von statistischen Begriffen und Modellen, deren Definitionen von evolutionären Prozessen inspiriert sind, **Optimierungsaufgaben** und andere Problemstellungen, die dann mittels MLA angenähert beziehungsweise bearbeitet werden.

Die Populationen des evolutionären Lernens bezeichnen verschiedene Suchräume, die das MLA im Laufe des Autoadaptionsprozesses bearbeitet. Die einzelnen Strukturvorschläge werden als **Individuen** bezeichnet. Evolutionäres Lernen kann die Leistungsfähigkeit einzelner Individuen direkt vergleichen und nutzt diese Möglichkeit, um evolutionäre Fortschritte beziehungsweise evolutionäre Veränderungen direkt an den Individuen festzumachen. In Konsequenz werden die Individuen untersucht und die Population als Gesamtobjekt wird nicht mit anderen Populationen verglichen, sondern bezeichnet jeweils lediglich die Gesamtheit aller zu einem gewissen Stand des Autoadaptionsprozesses gerade betrachteten Individuen. Die Population verändert sich bei evolutionärem Lernen in jedem Adaptionsschritt und die in iterativer Abfolge entstehenden Populationen werden als **Generationen** von Individuen bezeichnet. Populationen werden nur mit nachfolgenden Populationen verglichen und das auch nur, um den Fortschritt in der Entwicklung der Individuen zu veranschaulichen. Mit verschiedenen Individuen wird im maschinellen Lernen umgegangen, wie bei einer klassischen Betrachtung evolutionärer Prozesse mit verschiedenen

Populationen umgegangen würde. Die Begriffe sind im evolutionären Lernen jedoch nicht einfach vertauscht. Im Weiteren wird etwa dargestellt werden, wie Individuen im evolutionären Lernen gegeneinander antreten und dass sich einer der Gegner durchsetzt. Dieses Konzept lässt sich zwar prinzipiell, aber nicht ohne Weiteres auf Populationen übertragen. Wie bereits angedeutet wurde, stellt diese Verwendung der Begriffe in der Informatik kein Problem dar, da die MLA formal vollständig mit Mitteln der Informatik konstruiert und ausgewertet werden. Das maschinelle Lernen basiert meist nicht auf biologischen Modellen und es ist wichtig, dass Begriffe wie Evolution im Sinne der Informatik gelesen werden müssen, um einen interdisziplinären Zugang zu erhalten.

Evolutionäre Strategien lernen, indem sie ziellos experimentieren und erfolgreiche Ergebnisse weiterverfolgen. Gelernte Muster liegen in Form von bislang erfolgreichen Strukturvorschlägen vor. Das Zusammenspiel von Systematik und Zufall besteht bei evolutionärem Lernen darin, dass die Erzeugung der Designs für mögliche Strukturvorschläge un- beziehungsweise zufallsgesteuert ist, anschließend jedoch systematisch die besten Strukturvorschläge ausgewählt werden. Evolutionäres Lernen findet zufällig mögliche Strukturvorschläge und vergleicht diese systematisch. Dies wird im zweiten Hauptteil ein wesentlicher Schritt weg von Lernstrategien sein, die vollständig systematisch und auf Vorwissen basierend vorgehen beziehungsweise optimieren.

Evolutionäres Lernen eignet sich für den Einsatz in sehr komplexen Umgebungen, deren Hintergründe nicht verstanden werden. Diese Komplexität kann bewältigt werden, da nicht versucht wird, die Zusammenhänge der Umwelt zu erklären, sondern das System durch eine große Anzahl von zufälligen Veränderungen anzupassen.

Einführungsbeispiel zu evolutionärem Lernen

Als kurzes Anwendungsbeispiel soll der Entwurf eines elektronischen Schaltkreises dienen (Koza et al. 1996):

Im Vorfeld des eigentlichen Entwurfs werden die Anforderungen an das fertige Produkt formuliert und als **Spezifikationen** festgehalten. Weiterhin wird eine Simulationssoftware bereitgestellt, um Schaltkreisentwürfe auf ihre Leistungsfähigkeit zu testen. Schließlich werden alle für die Konstruktion des Schaltkreises verfügbaren Bauteile im Rahmen der Simulationssoftware als simulierter Werkzeugkasten dargestellt und es wird eine Anzahl simpler Standard-Schaltkreise entworfen, deren Gesamtheit als **Anfangspopulation** betrachtet wird. Sowohl der simulierte Werkzeugkasten als auch die Standard-Schaltkreise sind hierbei unabhängig von den genauen Spezifikationen und können unverändert von einem anderen Schaltkreisdiseign übernommen werden.

Nachdem diese Vorarbeit geleistet wurde, modifiziert ein evolutionär lernendes Artefakt die Individuen der Population – im ersten Schritt die Standard-Schaltkreise – indem zufällig Bauteile oder Verbindungen zwischen Bauteilen ergänzt oder entfernt werden. Anschließend vergleicht die Simulationssoftware die Ausgabewerte und ggf. die Konstruktionskosten der entstandenen Schaltkreise mit den geforderten Spezifikationen. Die leistungsstärksten Individuen werden ermittelt und als zweite Generation betrachtet. Die entstehenden Schaltkreise sind dabei in der Mehrzahl nicht funktionstüchtig, allerdings steigt der Anteil an funktionstüchtigen Schaltkreisen mit zunehmender Zahl von Generationen deutlich an. Das MLA erstellt solange weitere Generationen, bis die Spezifikationen ausreichend gut erfüllt werden.

Definition evolutionären Lernens

Auch im Allgemeinen sind vor dem Einsatz eines auf evolutionärem Lernen basierenden MLA zur Bearbeitung einer Problemstellung einige Vorarbeiten zu erledigen. Es wird eine Anfangspopulation benötigt und die enthaltenen Individuen werden meist als eine Zusammenstellung von auf Vorwissen beziehungsweise Rahmenbedingungen basierenden, unveränderlichen Einzelteilen modelliert, um zu vermeiden, dass zu viele nutzlose Lösungen entstehen. Zu diesem Zweck wird häufig eine **Repräsentation** der Problemstellung erstellt, die implizit den Suchraum der sinnvollen und damit syntaktisch zulässigen Strukturvorschläge vorgibt, woraus sich wiederum eine Anfangspopulation von Strukturvorschlägen gewinnen lässt. Diese Repräsentation kann auch eine **Codierung** sein, das heißt, die vorliegenden Rahmenbedingungen werden systematisch und wiederholbar in Eingabedaten übersetzt und so für das MLA registrierbar gemacht. Gleichzeitig unterstehen die aus der Codierung entstandenen Eingabedaten nach der Übergabe an das MLA nicht mehr den Gesetzmäßigkeiten, die außerhalb des MLA vorliegen. Der Verzicht auf eine Repräsentation im obigen Beispiel ist daran ersichtlich, dass die durch das MLA zurückgegebenen Strukturvorschläge nach einer **Decodierung**, das heißt einer Rückübersetzung in Schaltkreise gemäß derselben Systematik wie zuvor, zum Großteil nicht funktionstüchtig sind. Eine Codierung ist eine formale Grammatik mittels derer die Individuen zu beschreiben sind und entsprechend der alle Strukturvorschläge vom Nutzer interpretiert werden. Strukturvorschläge müssen entsprechend syntaktisch korrekt sein und die Codierung gibt damit den Autoadaptionsprozessen des evolutionären Lernens einen Rahmen. Die Codierung kann für einzelne Problemstellungen sehr unterschiedlich aussehen und die weiter unten besprochenen Varianten evolutionären Lernens unterscheiden sich insbesondere in den Methoden zur Konstruktion solcher einer Grammatik der Strukturvorschläge. In grober Analogie entspricht in der natürlichen Evolution die DNA solchen Codierungen, die nach Decodierung als Vorschlag für die Struktur eines Lebewesens interpretiert werden kann.

Im Autoadaptionsprozess des evolutionären Lernens werden die Anfangspopulation und jede darauffolgende Generation von Strukturvorschlägen durch an die Evolution angelehnte **evolutionäre Operatoren** so lange verändert, bis ein Strukturvorschlag entsteht, der nach einem vordefinierten Leistungsmaßstab hinreichend optimale Ergebnisse erzielt. Die typischen evolutionären Operatoren sind hierbei die zufällige **Mutation** eines Strukturvorschlags oder die **Rekombination** mehrerer Strukturvorschläge. Die Verwendung des Begriffs der Mutation bedeutet hier, dass ein kleiner Teil der Codierung zufällig abgeändert wird, während von Rekombination gesprochen wird, wenn syntaktisch vergleichbare Abschnitte – gegebenenfalls deutlich unterschiedlicher Länge – der Strukturvorschläge ausgetauscht werden. Die Bewertung der Individuen mittels des vordefinierten Leistungsmaßstabs wird als Einsatz der **Fitnessfunktion** und das Ergebnis der Bewertung als die **Fitness** der Individuen bezeichnet. Die Fitnessfunktion bewertet üblicherweise die Performanz über den Testdaten, kann aber auch Parameter wie die Komplexität des betrachteten Strukturvorschlags berücksichtigen. Evolutionäres Lernen kann somit als Optimierung bezüglich der Fitness verstanden werden.

Zusammengefasst werden die folgenden Prozessschritte durchlaufen:

A Initialisierung:

Der Prozess beginnt mit der Erstellung einer Codierung und der Zusammenstellung einer Anfangspopulation aus syntaktisch korrekt codierten Strukturvorschlägen.

B Evolutionsschritt:

Die aktuelle Generation wird mittels evolutionärer Operatoren evolviert.

C Selektionsschritt:

Die Strukturvorschläge der aktuellen Generation werden mittels einer vorgegebenen Fitnessfunktion beurteilt, und die fittesten Strukturvorschläge werden **probabilistisch** für die Erzeugung der nächsten Generation von Individuen, das heißt der nächsten Population, ausgewählt. Anschließend wird der nächste Evolutionsschritt **B** durchgeführt.

Probabilistisch bezeichnet die Zuordnung auf Basis einer Wahrscheinlichkeit und meint in der Praxis meist, dass die Wahrscheinlichkeit der Auswahl eines bestimmten Strukturvorschlags typischerweise dem Anteil der Fitness der entsprechenden Strukturvorschläge an der summierten Fitness der Population in der aktuellen Generation entspricht.

Sollte die Aufgabe darin bestehen spezielle geometrische Formen zu konstruieren, könnte eine Visualisierung des Evolutionsschrittes und Teilen des Selektionsschrittes wie folgt aussehen.

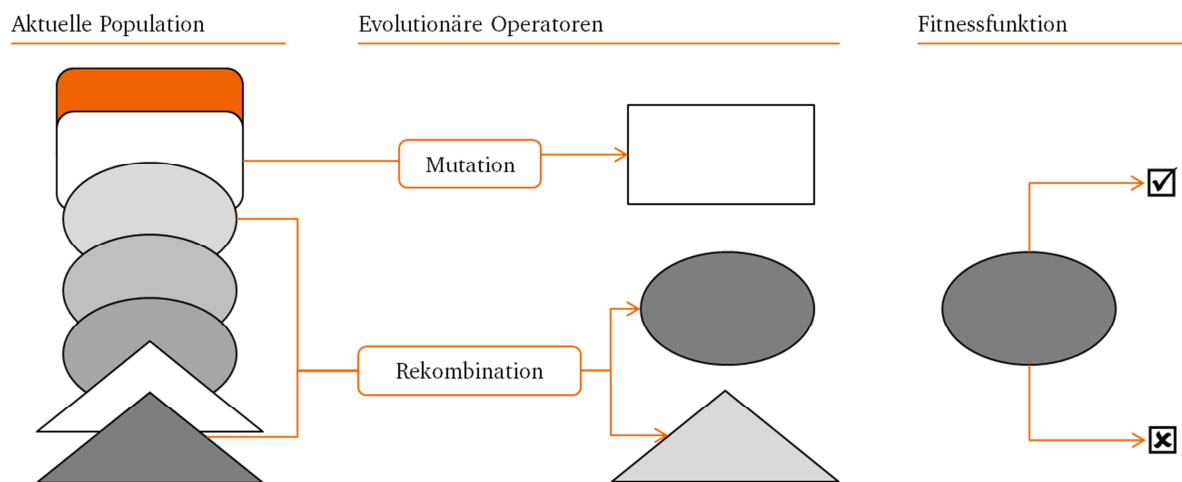


Abbildung 11: Evolutions- und Selektionsschritt am Beispiel geometrischer Objekte

Hier wurde mit Hilfe der Rekombination als evolutionärem Operator eine dunkelgraue Ellipse erzeugt. Diese Ellipse wird anschließend auf ihre Fitness überprüft, und sollte das Ziel beispielsweise darin liegen einen schwarzen Kreis darzustellen, so wäre die entstandene Ellipse das fitteste Individuum der derzeitigen Generation. In der Visualisierung fehlt für den Selektionsschritt noch die Erstellung einer neuen Generation.

Die Definition der gemeinsamen Grundlage der drei Klassen evolutionären Lernens hat eine noch nicht beleuchtete Besonderheit in Hinblick auf die Frage, inwieweit ein MLA im Rahmen des Autoadaptionsprozesses Sensordaten berücksichtigt oder berücksichtigen kann. Formal bestand dieses Problem schon bei der Betrachtung der MLA, die auf die Konstruktion von Entscheidungsbäumen aus waren. Allerdings wurde die Betrachtung dort ausgespart, da der Begriff der Fitnessfunktion die Analyse dringlicher macht und gleichzeitig erleichtert. Wie schon bei der Erstellung von Entscheidungsbäumen kann der gesamte Autoadaptionsvorgang ablaufen und einen Strukturvorschlag erzeugen, sobald die ursprünglichen Trainingsdaten übergeben wurden. Der resultierende Autoadaptionsprozess läuft bei der Konstruktion von Entscheidungsbäumen ohne spezielle Weiterentwicklungen immer gleich ab. Der Lernprozess bei evolutionärem Lernen kann jedoch eine autoadaptive Fitnessfunktion aufweisen.

Tatsächlich ist diese Option ein Hauptgrund dafür, dass die Fitnessfunktion einen eigenen Namen erhält und als prominenter Teil der Lernstrategie betrachtet wird. Im Falle des Einsatzes einer Fitnessfunktion, die sich im Laufe des Autoadaptionsprozesses verändert, hängt der von einem evolutionär lernenden MLA erzeugte Strukturvorschlag gegebenenfalls noch von weiteren Einflüssen wie der Reihenfolge der Eingabe der Trainingsdaten oder der Uhrzeit zu Beginn des Prozesses ab²⁹. MLA dieser Art könnten die Bewertung der Fitness einer Generation auch als Ausgabe an eine Testumgebung übergeben und als Eingabe die Bewertung der Fitness zurückerhalten. In diesem Fall wäre der Autoadaptionsprozess in gewisser Hinsicht immer nach einer Iteration beendet und andererseits würde er andauern und immer weitere Sensordaten aufnehmen, wenn sich die externen Anforderungen an die Fitness der Individuen häufig ändern. Wichtig ist hier zu verstehen, dass sowohl Entscheidungsbäume als auch evolutionäres Lernen am besten als einmal initiiertes abgeschlossenes System agieren können und die nachträgliche Integration von zusätzlichen Trainingsdaten zumindest kompliziert ist.

Die bei dieser Lernstrategie realisierte Suche unterscheidet sich durch ihre den Zufall nachbildenden Elemente deutlich von den Suchstrategien der übrigen MLA. Evolutionäres Lernen bekommt durch die Zufallselemente einen sehr unvorhersehbaren beziehungsweise unstetigen Charakter. Zwischen den Generationen können sehr unterschiedliche und zum Teil sehr extreme Veränderungen der Strukturvorschläge auftreten. Die Suche wird als **Strahlensuche** bezeichnet, bei der weder auf Basis eines groben Vorwissens gesucht, noch eine lokale Eigenschaft optimiert wird.

Vorteile und Nachteile evolutionären Lernens

Die Vorteile evolutionären Lernens entsprechen den Stärken, die auch der Evolution gemeinhin zugesprochen werden. MLA dieser Art sind eine robuste Methode der Anpassung an komplexe Systeme. Evolutionäres Lernen können Suchräume betrachten, die komplexe, aufeinander reagierende Elemente beschreiben, bei denen die Auswirkung einzelner Komponenten des Strukturvorschlages auf dessen Gesamtfitness schwierig zu modellieren ist. Beispiele für solche Suchräume sind die Optimierung einer Robotersteuerung oder eines hochmodularen Computerprogramms. Auch bei extrem großen Suchräumen kann evolutionäres Lernen aufgrund der Unvorhersehbarkeit der Suchschritte gute Ergebnisse erzielen. Analog zur Evolution benötigt evolutionäres Lernen keine oder kaum externe Steuerung und kann autoadaptiv konstruiert werden. Das bedeutet, die Eigenschaften der evolutionären Operatoren und der meisten sonstigen Parameter der Lernstrategien können dynamisch angepasst werden. Der Verzicht auf eine externe Steuerung führt nicht zwangsläufig dazu, dass das System ziellos agiert, allerdings besteht sehr wohl die Möglichkeit auf ein explizites Ziel zu verzichten und lediglich eine rudimentäre Vorgabe für eine zwar ziellose, aber dennoch systematische Autoadaptation zu machen. Wird diese Vorgabe zusätzlich autoadaptiv gestaltet, so können zuvor nicht bedachte Muster gefunden werden, allerdings senkt diese Vorgehensweise aufgrund der wenig vorstrukturierten Suche stark die Geschwindigkeit des MLA. Aus praktischer Sicht schließlich zeichnet sich evolutionäres Lernen dadurch aus, dass es sich einfach parallelisieren lässt und damit sehr effizient implementiert werden kann.

²⁹ Ein Algorithmus kann keine echten Zufallsgrößen erzeugen. Zufallszahlengeneratoren nutzen häufig quasi-zufällige Parameter wie die Systemzeit um annähernd zufällige Ergebnisse zu erzeugen.

Von Nachteil bei evolutionärem Lernen ist vor allem, dass es relativ zu anderen Lernstrategien in unmodifizierter Form – analog zu biologischer Evolution – sehr langsam seine Performanz über den Trainingsdaten verbessern und dass es schwieriger ist, eine Erfolgsgarantie für eine zufallsgeprägte Suche zu erstellen beziehungsweise zu errechnen. Weiterhin ist die Wahl der Codierung der Strukturvorschläge von sehr großer Bedeutung, da die Suche syntaktische Einschränkungen bei der Formulierung der Anfangspopulation mitunter nicht überwinden kann.

Varianten evolutionären Lernens

Evolutionäres Lernen hängt, wie bereits angedeutet, sehr stark von der Form der gewählten Modellierung beziehungsweise Codierung der Problemstellung ab. Die Zahl der Abhängigkeiten ist jedoch noch deutlich größer, etwa üben auch Aspekte wie die Auswahl der einzusetzenden evolutionären Operatoren einen großen Einfluss aus, und sogar scheinbar nachrangige Parameter wie die Anwendungsreihenfolge spielen eine Rolle. Diese Vielzahl von Einflussfaktoren ist insofern bemerkenswert, als evolutionär lernende Artefakte durch Manipulation dieser Faktoren sehr unterschiedliche Vorgehensweisen aufweisen können. Die drei wichtigsten Formen evolutionären Lernens lassen sich darüber hinaus gut interdisziplinär betrachten. Unabhängig von den Spezifika evolutionären Lernens bietet sich so die Möglichkeit, ein Verständnis für das Zusammenspiel unterschiedlicher Varianten eines gemeinsamen Ansatzes maschinellen Lernens zu erlangen. Solch ein Verständnis ist auch in der weiteren Analyse des maschinellen Lernens hilfreich.

Die drei wichtigsten Konkretisierungen der Idee des evolutionären Lernens und die zugrundeliegenden Ideen werden im Weiteren kurz vorgestellt. Diese wesentlichen Varianten sind:

- Genetische Algorithmen
- Genetische Programmierung
- Evolutionsstrategien

Ergänzend zu den hier vorgestellten Varianten evolutionären Lernens können Hypothesen auch durch symbolische Repräsentationen beschrieben werden – wie etwa im Beispiel der geometrischen Figuren in Abbildung 11.

Genetische Algorithmen

Codierung im Rahmen genetischer Algorithmen

Beim Einsatz **genetischer Algorithmen** werden Strukturvorschläge typischerweise als **Bitfolgen** – Ketten von Nullen und Einsen – codiert, die im Kontext des Problems interpretiert werden müssen. Dies bedeutet, dass die Attribute, die bei der Codierung eines Strukturvorschlages von Bedeutung sind, auf eine festgelegte Weise oder an einer festgelegten Stelle in der Bitfolge hinterlegt sind und dass diese Codierung in ihrer Gesamtheit den vorliegenden Strukturvorschlag beschreibt.

Der Rolle der Bitfolge würde bei einem Lebewesen in etwa die Rolle der DNA entsprechen. Die Codierung mittels Bitfolgen wird entsprechend als **Genotyp-Phänotyp-Abbildung** von Bitfolgen auf die Menge der Strukturvorschläge betrachtet werden. Die wesentliche Idee hinter dieser Form der Codierung liegt darin, dass eine möglichst **atomare**, das heißt minimal komplexe, Beschreibungssprache eine maximale Ausdrucksfähigkeit erzeugt (Goldberg 1990,

S. 4ff). Bitfolgen eignen sich sehr gut für die Codierung von Kausalsätzen, da jedes Bit als Entscheidung für oder gegen etwas interpretiert werden kann. Ein konkretes Beispiel für die Codierung eines Strukturvorschlages ist etwa das Problem, zu entscheiden, ob eine Kanufahrt unternommen werden soll.

Problemstellung	Beobachtbare Größe:
Kanu fahren: <input checked="" type="checkbox"/> <input type="checkbox"/>	Wetter: + 0 -

Abbildung 12: Grundlage einer Beispielcodierung von „Kanu fahren“

In diesem Beispiel ist das Wetter das einzige Attribut und kann die drei Zustände gutes, normales und schlechtes Wetter annehmen. Ein Strukturvorschlag, der in jeder Situation eine Entscheidung ermöglicht, muss jeden möglichen Zustand des Attributes berücksichtigen. Ein möglicher Strukturvorschlag in Form eines Kausalsatzes wäre die Formulierung „gutes und normales Wetter sind akzeptabel, schlechtes Wetter ist es nicht“. Zu drei möglichen Zuständen muss jeweils eine Empfehlung ausgesprochen werden, deshalb könnte für die Codierung des Kausalsatzes ein dreistelliger Bitstring eingesetzt werden. Das genannte Beispiel „ja-ja-nein“ als ein möglicher Strukturvorschlag könnte mit „110“ in Form einer Bitfolge codiert werden. Wenn ein zweites Attribut betrachtet werden soll, könnte die Bitfolge einfach um so viele Stellen verlängert werden, wie das neue Attribut Zustände einnehmen kann. Die ersten drei Stellen der Bitfolge wären anschließend für die bisherige Codierung reserviert und die restlichen Stellen für die Codierung der Aussage des Strukturvorschlages bezüglich des zweiten Attributes. Wenn in einem Strukturvorschlag ein einzelnes Attribut keine Rolle spielt, können die reservierten Stellen pauschal mit einer eins beschrieben werden. Wenn etwa in obigem Beispiel das Wetter irrelevant ist und die Kanufahrt unabhängig davon sowieso stattfinden wird, entspräche das dem Strukturvorschlag „111“. Diese Vorgehensweise stellt sicher, dass die Bitfolgen aller Strukturvorschläge die gleiche Codierungslänge aufweisen. Dadurch wird es deutlich einfacher sicherzustellen, dass jede eingesetzte Bitfolge auch einen syntaktisch korrekten Strukturvorschlag darstellt³⁰.

Evolutionäre Operatoren und Selektion bei genetischen Algorithmen

Genetische Algorithmen nutzen sowohl Mutation als auch Rekombination als evolutionäre Operatoren, wobei der Schwerpunkt häufig auf dem Einsatz der Rekombination liegt. Mutationen werden realisiert, indem ein zufälliges Bit geändert wird. Rekombinationen werden durchgeführt, indem Teile von zwei Bitfolgen ausgetauscht werden. Es wird nach einem vorgegebenen oder zufälligen Muster entschieden, welche Teilstücke zwischen den beiden Bitfolgen ausgetauscht werden, und zwei neue Bitfolgen mit ausgetauschten Teilen werden erzeugt. Das vorgegebene Muster zu Rekombination wird **Maske** genannt.

Genetische Algorithmen können in einem gewissen Rahmen ihre eigenen evolutionären Operatoren adaptieren, indem sie beispielsweise die codierte Maske explizit als einen zusätzlichen Teil der eigentlichen Bitfolge des Strukturvorschlages betrachten und evolvierten. Auf diese Weise können Entscheidungen über die Häufigkeit des Einsatzes von evolutionären

³⁰ Genetische Algorithmen können auch auf Basis von Eingaben arbeiten, die nicht in Form einer Bitfolge vorliegen. Entsprechende Varianten genetischer Algorithmen, die andere Formen einer Genotyp-Phänotyp Analogie aufweisen, werden in der Praxis auch erfolgreich umgesetzt (Salomon 1995).

Operatoren bei späteren Generationen getroffen werden. Darüber hinaus können die evolutionären Operatoren selbst verändert werden, etwa indem die Maske als Teil des Strukturvorschlags der letzten Generation ebenfalls mutiert. Dieses Vorgehen ermöglicht genetischen Algorithmen zumindest prinzipiell mittels Autoadaptation zu neuen Suchstrategien zu gelangen. Dies ist theoretisch auch bei den anderen Formen evolutionären Lernens umsetzbar, wird dort jedoch kaum genutzt.

Der Selektionsschritt bei genetischen Algorithmen entspricht dem Grundmuster für evolutionäres Lernen. Die Selektion wird auf Basis einer Fitnessfunktion vorgenommen und nur die fittesten der neu entstandenen Strukturvorschläge werden Teil der nächsten Generation von Individuen.

Genetische Programmierung

Codierung im Rahmen genetischer Programmierung

Der Ansatz der **genetischen Programmierung** liegt darin, die Auswirkung evolutionärer Operatoren auf Algorithmen zu betrachten. Algorithmen können sehr unterschiedlich repräsentiert werden, werden im Kontext der genetischen Programmierung aber häufig als **Bäume** dargestellt. Der sehr kurze Algorithmus zur Auswertung der Formel „A + 2B“ kann etwa in der folgenden Form als Baum dargestellt werden.

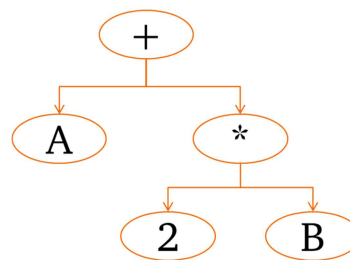


Abbildung 13: Auswertungsbaum beziehungsweise Parse Tree der Formel „A + 2B“

Die Idee hinter der genetischen Programmierung ist, dass keine zusätzliche Codierung vorgenommen werden muss, falls die Individuen der aktuellen Population ein Zusammenwirken von Algorithmen darstellen. Auch die Repräsentation als Baum stellt keinen zusätzlichen Schritt dar, da Algorithmen von der ausführenden Hardware meist sowieso in solch einer Form abgearbeitet werden³¹. Auf Basis dieser Repräsentation können ganze Algorithmen evolviert werden. Wenn keine zusätzliche Codierung notwendig ist, weil die Strukturvorschläge bereits codiert vorliegen, unterscheiden die Einwirkungen von evolutionären Operatoren auf die Codierung sich nicht von Einwirkungen direkt auf den Strukturvorschlag selbst. Die Unterscheidung zwischen **Genotyp** und **Phänotyp**³² des Individuums, die bei genetischen Algorithmen zentral war, wird irrelevant. Die evolutionären Operatoren können stattdessen direkt auf den Phänotyp zugreifen. Dies hat den Nachteil, dass, auch wenn das Evolvieren eines Algorithmus prinzipiell jede denkbare oder sinnvolle Form erzeugen kann, dies in der Umsetzung sehr komplexe oder schlicht sehr lange

³¹ Die Darstellung der Strukturvorschläge als Baum entspricht einem **Parse Tree** oder „Syntaxbaum“ des gewählten Algorithmus, wie in der obigen Abbildung bereits angedeutet wurde.

³² Der Genotyp ist die genetische Codierung, das heißt die Information der Gene, die eine biologische Zelle im Zellkern trägt. Der Phänotyp stellt die Realisierung dieser Codierung dar – etwa in Form einer Haarfarbe.

Strukturvorschläge erfordern wird. Ein Beispiel hierfür aus dem Schachspiel ist der Versuch, bestimmte Felder mit Hilfe einer speziellen Figur zu erreichen. Ein Strukturvorschlag, der eine Springer-Zugfolge beschreibt, ist in den meisten Fällen deutlich länger als ein Strukturvorschlag für den Einsatz einer Dame. Analog kann im Rahmen der genetischen Programmierung auf sehr unterschiedliche Hilfsalgorithmen zurückgegriffen werden. Die Hauptschwierigkeit beim Einsatz genetischer Programmierung besteht entsprechend darin diejenigen Hilfsalgorithmen zu identifizieren, die zur Bearbeitung der konkreten Problemstellung besonders geeignet sind. Die Identifizierung eines im jeweiligen Kontext gut einsetzbaren Algorithmus führt meist überhaupt erst zur Wahl der genetischen Programmierung als Lernstrategie. In Konsequenz verschiebt die genetische Programmierung das Problem der Wahl eines geeigneten Algorithmus zunächst nur. Allerdings bieten sich durch den bei der Verschiebung gewonnenen neuen Kontext auch neue Ansätze Strukturvorschläge zu erstellen.

Evolutionäre Operatoren und Selektion bei genetischer Programmierung

Die Verwendung von evolutionären Operatoren bei der genetischen Programmierung ist nicht eindeutig zu beschreiben. Sowohl die Rekombination von Teilbäumen in Form des Austauschs zweier Äste als auch die Mutation einzelner Knoten werden eingesetzt.

Die Selektion erfolgt bei genetischer Programmierung prinzipiell analog zu derjenigen bei genetischen Algorithmen, allerdings erfolgt die Auswahl häufig als **Turnierselektion** oder **Rouletteleselektion**. Die Turnierselektion lässt die für die Aufnahme in die nächste Generation in Frage kommenden Individuen in Form eines Turnieres gegeneinander antreten, wobei jeweils das Individuum mit der niedrigeren Fitness ausscheidet. Die bestplatzierten Individuen werden anschließend in die nächste Generation aufgenommen. Wenn die Paarung von Individuen zufällig war, kann so durchaus das zweitbeste Individuum im Selektionsschritt ausgesondert werden und entsprechend ist eine gewisse Durchmischung jenseits der Fitness gewährleistet. Die Rouletteleselektion ordnet jedem Individuum eine von dessen Fitness abhängige Wahrscheinlichkeit zu, mit der das Individuum in die nächste Generation aufgenommen wird. Die Wahrscheinlichkeit kann sich dabei an der absoluten Höhe der Fitness orientieren und Individuen mit sehr viel höherer Fitness auch sehr viel höhere Wahrscheinlichkeiten zuordnen, oder die Fitness relativ zu den anderen Individuen wird als Platzierung interpretiert. Im zweiten Fall erhält das Individuum mit der höchsten Fitness die größte Wahrscheinlichkeit und das Individuum mit der geringsten Fitness die niedrigste Wahrscheinlichkeit zur Aufnahme in die nächste Generation. Beide Selektionsweisen sollen vermeiden, dass die genetische Programmierung zu einer reinen Optimierung der Fitness wird. Genetische Algorithmen setzen zu diesem Zweck autoadaptive genetische Operatoren ein, während genetische Programmierung einen zufälligen Aspekt in den Selektionsschritt aufnimmt. Dieser Versuch der aktiven Distanzierung von **zielorientierten Optimierungsalgorithmen** deutet an, dass die später im zweiten Hauptteil vorgeschlagene Unterscheidung optimierungsnäherer und -fernerer Ausprägungen maschinellen Lernens auch für die Informatik interessant sein kann.

Beispiele für genetische Programmierung

Ein Beispiel für genetische Programmierung bildet die Suche nach einem Strukturvorschlag zur Lösung des Spiels „Turmbau zu Hanoi“.

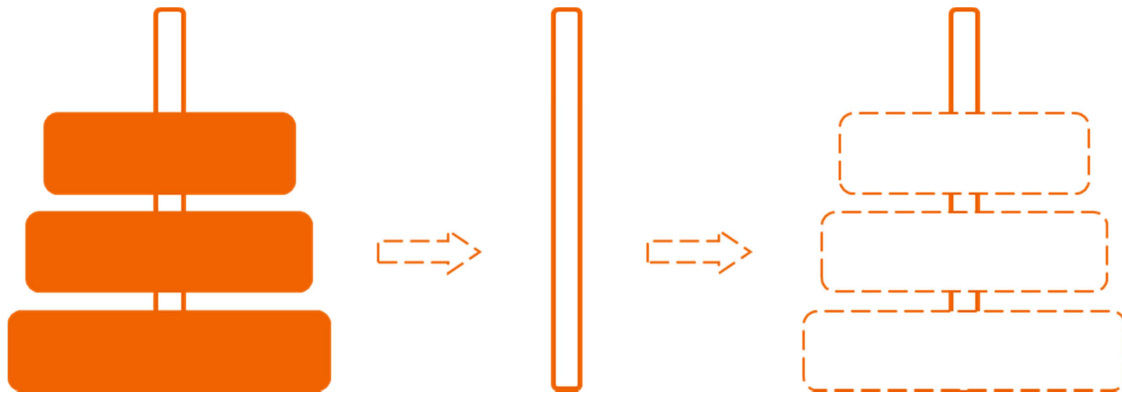


Abbildung 14: Aufbau und Zielzustand des Turmbaus von Hanoi

Die Aufgabenstellung beim Turmbau von Hanoi beginnt mit der Vorgabe von drei Pfählen, wobei auf einem Pfahl drei der Größe nach sortierte, unterschiedlich breite Scheiben platziert sind. Ziel ist es, diese Scheiben in der gleichen Anordnung auf einem anderen Pfahl zu platzieren. In jedem Zug darf immer nur genau eine Scheibe auf einmal bewegt werden und es dürfen auf allen Pfählen nur kleinere auf größeren Scheiben platziert werden dürfen. Die denkbaren Züge entsprechen bei dieser Aufgabe den Bewegungen einer speziellen Scheibe von einem bestimmten Pfahl zu einem anderen Pfahl und sind algorithmisch einfach beschreibbar. Strukturvorschläge setzen sich direkt aus einer Anreihung von Zügen zusammen, und sowohl die Zulässigkeit der Zugfolge als auch deren Funktionstüchtigkeit sind einfach zu simulieren und direkt durch den Nutzer überprüfbar. Wesentlich ist hier, dass die denkbaren Züge als fixe Komponenten des Strukturvorschlages betrachtet werden. Die Menge der denkbaren Züge darf nicht erweitert werden und nur die Reihenfolge der Züge darf durch evolutionäre Operatoren evolviert werden. Die Art und Weise, wie die Reihenfolge evolviert wird, darf wiederum selbst evolviert werden, das heißt die evolutionären Operatoren können autoadaptiv sein. Gleichzeitig können durch die Fixierung der denkbaren Züge dennoch keine überraschenden Lösungen auftreten. Denkbar wären durchaus auch evolvierte Zugoptionen, etwa könnte die Hälfte der dritten Scheibe auf den zweiten Pfahl geschoben werden. Wenn in der Praxis alle undenkbbaren Züge tatsächlich auch nicht umsetzbar sind, ist solch eine Vorgehensweise wenig hilfreich, aber in der Praxis treten selten unveränderliche Rahmenbedingungen auf, gegen die nicht verstoßen werden darf. Ein MLA kann hier helfen, die Menge der denkbaren Reaktionen im Kontext einer speziellen Aufgabe nur als einen Teil des Suchraums für den Strukturvorschlag zu betrachten.

Ein etwas komplexeres Beispiel für genetische Programmierung ist das eingangs beschriebene automatisierte Design von elektronischen Schaltkreisen (Koza et al. 1996), das in der Praxis mittels genetischer Programmierung realisiert werden kann. Die technischen Details einer solchen Realisierung stehen in der interdisziplinären Betrachtung des maschinellen Lernens nicht im Fokus, allerdings ist es hilfreich eine Intuition der Größenverhältnisse der Parameter zu entwickeln. Ergänzend zur eingangs erfolgten Darstellung daher nachfolgend Größenordnungen der Parameter aus einer praktischen Anwendung:

- In jeder Generation bestand die Population aus über 500.000 Schaltkreisentwürfen.
- Die besten 10% jeder Generation wurden unverändert wiederverwendet, um mehr Mutationen und Rekombinationen zu erlauben, da bereits eine große Anzahl von performanten beziehungsweise fitten Individuen aus der aktuellen Generation übernommen und somit ein größeres Risiko eingegangen werden konnte.

- 1% der Individuen der nächsten Generation wurde durch Mutationen, der übrige Anteil durch Rekombinationen gewonnen.
- Der Anteil der Schaltkreisentwürfe, die im Rahmen der Simulationssoftware sinnvoll dargestellt werden konnten, begann bei weniger als 5% und stieg im Laufe der nächsten Generationen erst auf 15%, dann auf 25%. Im Schnitt über alle Generationen entstanden etwa zu 90% sinnvolle Ergebnisse.
- Nach ungefähr 140 Generationen entstand ein Schaltkreis mit den gewünschten Spezifikationen.

Der vergleichsweise seltene Einsatz von Mutationen überrascht auf den ersten Blick, allerdings gilt hier dieselbe implizite Annahme wie beim Turmbau von Hanoi. Die Schaltkreiselemente, die dem MLA als mögliche Komponenten für die Erstellung eines Strukturvorschlages genannt wurden, entsprechen den denkbaren Lösungen. Zwar ist beim Schaltkreisdesign klar, dass auch sehr nützliche Designs noch nicht entdeckt wurden, allerdings wird der Bereich von den Nutzern als gut verstanden wahrgenommen, und die Chancen auf halb-zufällige Entdeckungen durch ein MLA werden als fast vernachlässigbar erachtet. In anderen Kontexten kann sich diese Einschätzung natürlich drastisch ändern.

Evolutionsstrategien

Die Grundannahme und Voraussetzung für die Verwendung von Evolutionsstrategien³³ ist, dass im jeweiligen Kontext kleine Änderungen der Strukturvorschläge nur kleine Änderungen in deren Performanz und anderen wichtigen Eigenschaften herbeiführen. Dies entspricht der Annahme einer hinreichend **starken Kausalität** als universellem Weltverhalten. Die zur Performanzmessung verwendete Fitnessfunktion wird bei Evolutionsstrategien als **Zielfunktion** bezeichnet. Das ist insofern von Bedeutung, als der Begriff der Zielfunktion sich üblicherweise in der **mathematischen Optimierung** wiederfindet, in der versucht wird eine Zielfunktion zu **maximieren**, indem Parameter verändert werden, die die Zielfunktion beeinflussen.

Die Annahme einer starken Kausalität soll sicherstellen, dass beim Einsatz der evolutionären Operatoren Mutationen, die zu besonders großen Veränderungen im Strukturvorschlag oder dessen Eigenschaften führen, und Mutationen, die zu sehr kleinen Veränderungen führen, langfristig systematisch mit geringerer Wahrscheinlichkeit zu einem Fortschritt im Sinne der Zielfunktion führen als Mutationen, die eine zu bestimmende, optimale Größe von Veränderungen herbeiführen. Das Ausmaß der Veränderungen, die eine Mutation auslöst, wird als **Mutationsschrittweite** bezeichnet. Ein Ziel von Evolutionsstrategien besteht darin **Mutationsschrittweitenband** zu bestimmen, das aussagt in welchem Rahmen die Mutationsschrittweite in einem Evolutionsschritt minimal und maximal liegen sollte. Entsprechend ist ein zentrales Element von Evolutionsstrategien die Adaption der Mutationsschrittweite und daraus wiederum folgt, dass Evolutionsstrategien im Gegensatz zu anderen Formen evolutionären Lernens bereits in ihrer Grundform autoadaptive evolutionäre Operatoren einsetzen. Evolutionsstrategien besitzen darüber hinaus aufgrund der Voraussetzung einer starken Kausalität und der gezielten Steuerung der Mutationsschrittweite im Gegensatz zu den anderen Formen evolutionären Lernens ein mathematisches nutzbares Fundament. Auf dieser Basis können mathematische Analysen der **Konvergenz** des

³³ Der Begriff der Evolutionsstrategie wird mitunter in der Informatik unterschiedlich verwendet, allerdings ist die hier beschriebene Verwendung sehr häufig anzutreffen.

Autoadaptionsprozesses durchgeführt werden, das heißt Aussagen über die Existenz und Beschaffenheit eines Strukturvorschlages maximaler Fitness werden möglich.

Codierungen bei Evolutionsstrategien

In Rahmen von Evolutionsstrategien wird ein Individuum durch eine Menge von individuellen **Objektparametern** mit zugewiesenen Zielfunktionswerten codiert. Realisiert wird diese Codierung der Strukturvorschläge durch einen **Vektor**, bei dem Nachkommastellen zugelassen sind. Vektoren sind mathematische Darstellungsformen für angeordnete Werte und werden in folgender Notation dargestellt.

$$\begin{pmatrix} 2,5 \\ 1 \\ 0 \\ 3,4 \end{pmatrix}$$

Abbildung 15: Notationsbeispiel eines Vektors mit zufälligen Werten

Die Verwendung von Nachkommastellen ermöglicht unendlich viele verschiedene Eingabemöglichkeiten, etwa 2,1 oder 2,11 oder 2,111 und so fort. Im Gegensatz dazu codierten etwa die Bitfolgen genetischer Algorithmen typischerweise eine endliche Zahl von möglichen Zuständen.

Evolutionäre Operatoren und Selektion bei Evolutionsstrategien

Wie die genetische Programmierung versuchen auch Evolutionsstrategien eine komplexe oder undurchsichtige Codierung zu vermeiden und sprechen nicht von einer Unterscheidung zwischen einem Genotyp und einem Phänotyp der Strukturvorschläge. Allerdings verwenden auch Evolutionsstrategien Begriffe der Biologie, um dem Verständnis der Algorithmen zuträglich Assoziationen zu erzeugen. Evolutionsstrategien sprechen von selbstadaptiven³⁴ und fixen **Strategieparametern**, die im selbstadaptiven Fall als **endogen** und anderenfalls **exogen** bezeichnet werden. Die Unterscheidung soll andeuten, dass exogene Strategieparameter dem Kontext der Aufgabe entstammen, die vom jeweiligen MLA bearbeitet werden soll, und nicht evolviert werden können. Die Strategieparameter beziehen sich auf die Strategie bei dem Einsatz eines evolutionären Operators und sind im Einzelnen:

- Die Größe der zu evolvierenden Population
- Die Anzahl der Elternindividuen, die **multi-rekombinativ** Nachkommen generieren
- Die Anzahl der von jeder Gruppe von Eltern erzeugten Nachkommen
- Die Entscheidung, ob Elternindividuen bei der nächsten Selektion mitberücksichtigt werden

Die Anzahl der Elternindividuen wird als **Mischungszahl** bezeichnet und liegt beim Menschen bei zwei Eltern. Evolutionsstrategien verwenden ebenso wie das übrige evolutionäre Lernen Mutation und Rekombination als evolutionäre Operatoren, denn wenn die Mischungszahl eins beträgt, entspricht dies einer Mutation, da genau ein **Elter** evolviert wird und eine festzulegende Anzahl Nachkommen generiert wird. Durch die Formalisierung des Evolutions-

³⁴ „Selbstadaptiv“ ist dabei synonym zu der Rede von Autoadaptivität in dieser Arbeit zu verstehen.

schrittes ist ein großes Spektrum von evolutionären Operatoren darstellbar, die prinzipiell auch bei anderen Formen evolutionären Lernens Verwendung finden könnten. Hierin findet sich auch die Motivation der Rede von evolutionären Operatoren, die bisher nur Mutationen und Rekombinationen unter einem Begriff zusammenfassen konnte. Jetzt sind Mutationen nur spezielle **Parametrisierungen** allgemeiner Strategieparameter. Die zentrale Rolle der Suche nach dem Mutationsschrittweitenband führt bei Evolutionsstrategien im Vergleich zu anderen Formen evolutionären Lernens zu einem Schwerpunkt auf der Mutation gegenüber der Rekombination. Dieser Schwerpunkt ist relativ, da die Mutation häufig deutlich mehr Relevanz erhält als bei anderen Formen evolutionären Lernens, allerdings dadurch innerhalb der Evolutionsstrategie nicht zwangsläufig häufiger eingesetzt wird als die Rekombination. Insbesondere gibt es sehr viel mehr Möglichkeiten zur Rekombination als zur Mutation, da der entsprechende Strategieparameter nur in genau einem Fall eine Mutation erzeugt und in allen anderen Fällen Rekombinationen erstellt – auch und gerade bei sonst identischen Strategieparametern. Ein Ausdruck der Betonung von Mutationen ist, dass Evolutionsstrategien mitunter systematisch erst Rekombinationen durchführen und Mutationen dadurch auf einer größeren Menge von Individuen durchgeführt werden können.

Evolutionsstrategien bestimmen durch eine Parameterwahl, ob Elternindividuen in der Selektion berücksichtigt werden sollen, dadurch entsteht eine Verbindung des Evolutionsschrittes mit dem Selektionsschritt. Die Selektion findet als abschließender Teil des Evolutionsschrittes statt und die Wahl, welche Individuen der letzten Generation evolviert werden, wird vor Durchführung der Selektion getroffen. Andere Formen evolutionären Lernens trennen den Evolutionsschritt und den Selektionsschritt strikt. Lernstrategien, bei denen nur diejenigen Individuen weiter betrachtet werden, die eine hohe Fitness aufweisen, können so gedacht werden, dass sie jeden Autoadaptionszyklus mit dem Selektionsschritt beginnen. Diese Denkweise bestärkt die Wahrnehmung, dass zuerst die Fitness der letzten Generation bewertet wird, bevor festgelegt wird, wie die neue Generation sich zusammensetzt, dass also Evolutionsschritt und Selektionsschritt strikt getrennt sind. Die unterschiedlichen Interpretationen lassen sich wie folgt visualisieren.

1. Evolution, 2. Selektion → 1. Evolution, 2. Selektion → 1. Evolution, 2. Selektion

1. Selektion, 2. Evolution → 1. Selektion, 2. Evolution → 1. Selektion, 2. Evolution

Abbildung 16: Unterschiedliche Autoadaptionszyklen im evolutionären Lernen

Schon bei der Definition evolutionären Lernens deuteten sich diese beiden Interpretationsmöglichkeiten an, dort wurde eine einmalige Initialisierung durchgeführt, die als eine erste Selektion interpretiert werden kann.

Problembehandlungen und Weiterentwicklungen evolutionären Lernens

Zur Vermeidung von Überanpassung können innerhalb eines Strukturvorschlages analog zum Stutzen von Entscheidungsbäumen zufällig gewählte Einschränkungen bezüglich eines Attributes getilgt oder gleich alle Anforderungen an ein bestimmtes Attribut aus dem Strukturvorschlag entfernt werden.

Evolutionäres Lernen tendiert zu **Crowding**, womit das Auftreten einer Gruppe von Individuen bezeichnet wird, die sich untereinander sehr ähneln und ein höheres Maß an

Fitness aufweisen als die übrige Population. Crowding ist selbstverstärkend, da in den nächsten Evolutionsschritten wiederum sehr viele Mutationen und Rekombinationen der fittesten Individuen entstehen. Eine Maßnahme zur Vermeidung dieses Phänomens ist der Einsatz von Turnier- oder Rouletteselektion, eine andere Maßnahme besteht darin zu fordern, dass jedes Individuum nur einmal als Elter zum Einsatz kommt oder dass die Anzahl an Nachkommen pro Individuum begrenzt wird. Weiterhin kann die gemeinsame Elternchaft von sehr ähnlichen Individuen verboten oder erzwungen werden um Crowding zu erzeugen oder zu vermeiden.

Eine dritte typische Weiterentwicklung evolutionären Lernens besteht in der Nutzung des **Baldwin Effekts**. Der Baldwin Effekt beschreibt, dass Individuen in einer sich verändernden Umwelt einen evolutionären Vorteil besitzen, falls sie in der Lage sind, unabhängig von der Entwicklung der Population, zu der sie gehören, lernen zu können, das heißt in der Lage sind mittels lokaler Autoadaptionsprozesse individuell ihre Fitness zu erhöhen. Individuen, denen individuelles Lernen erlaubt ist, müssen weniger gut an spezifische Situationen angepasst sein. Die Nutzung dieses Effektes kann im Rahmen der Wahl der Strategieparameter realisiert werden, indem evolutionäre Operatoren gezielt auf Teilpopulationen angewendet werden, die etwa bezüglich der Fitnessfunktion eine gemeinsame Schwäche zeigen. Wenn Schwächen von Teilpopulationen auf diese Weise gezielt reduziert werden, kann die Population insgesamt gegebenenfalls in größeren Schritten mutieren und das Mutationsschrittweitenband erweitert sich oder kann sogar hin zu größeren Schrittweiten verschoben werden.

Viele Ideen und Problemlösungen des evolutionären Lernens sind aus der Biologie motiviert und an real existierende Phänomene angelehnt. Das soll nicht darüber hinwegtäuschen, dass die Vorgehensweise häufig diejenige einer mathematischen Optimierung ist und der Zufallsfaktor und die Ziellosigkeit der natürlichen Evolution keine unmittelbare Entsprechung haben. Wichtig für die Diskussion des zweiten Hauptteils wird aber sein, dass innerhalb der Informatik im Zusammenhang mit maschinellem Lernen Maßnahmen entwickelt werden, mit Hilfe derer solch eine zufällige Ziellosigkeit nachempfunden und in MLA erzeugt werden kann.

2.3.4. Lernen von künstlichen neuronalen Netzen – KNN

Motivation

Künstliche neuronale Netze – kurz **KNN** – bilden ein Feld des maschinellen Lernens, das sich für den Entwurf eines Autoadaptionsprozesses die Prozesse und Strukturen innerhalb des menschlichen Gehirns zum Vorbild genommen hat. Analog zu der Motivation des evolutionären Lernens bezieht sich auch bei KNN die Motivation auf vereinfachende Aussagen zur Funktionsweise des Gehirns. Ein Beispiel für solch eine Aussage ist die 100-Schritt-Regel.

Ein Mensch kann einen ihm bekannten Gegenstand oder eine bekannte Person innerhalb von 0,1 Sekunden erkennen. Dabei sind bei einer angenommenen Verarbeitungszeit einer Nervenzelle von 1 Millisekunde maximal 100 sequentielle Verarbeitungsschritte im Gehirn des Menschen nötig.

(Wikipedia Contributors 2012, 100-Schritt-Regel)

Diese Leistungsparameter werden von technischen Systemen zur Objekterkennung noch nicht erreicht. Das Gehirn scheint im Gegensatz zu modernen Rechnern zu einer massiven und funktionellen **Parallelverarbeitung**³⁵ in der Lage zu sein.

Der Fokus beim Versuch der Übertragung dieser Fähigkeit auf MLA liegt nicht auf einer präzisen Modellierung der extrem komplizierten biochemischen Vorgänge des Gehirns. Stattdessen wird die Betrachtung auf einen verhältnismäßig gut verstandenen Teilbereich dieser Prozesse beschränkt, auf die Funktionsweise spezieller **Nervenzellen** im Gehirn, der **Neuronen**. Künstliche neuronale Netze basieren, genau wie der Name es andeutet, auf der Betrachtung von vernetzten künstlichen Neuronen. Weitere Eigenschaften des Gehirns, wie eine hohe Parallelität von Prozessen, werden von KNN genau dann genutzt, wenn diese sich gut in den Methoden der Informatik nutzen lassen. Andere Eigenschaften wie hormonelle Abhängigkeiten werden nicht betrachtet.

Kurzgefasst liegt die Motivation für die Nutzung von KNN in der Hoffnung, die Rolle der Neuronen für die biologischen Vorgänge im Gehirn imitieren und mittels künstlicher Neuronen die Stärken neuronaler Strukturen auf MLA übertragen zu können. Autoadaptionsprozesse und die resultierenden Strukturvorschläge auf Basis von KNN können auch tatsächlich trotz des stark vereinfachenden Vorgehens erfolgreich einige Stärken des menschlichen Gehirns reproduzieren, etwa eine große Unempfindlichkeit gegenüber verfälschten und unvollständigen Eingaben.

Abgrenzung zu biologischen Gehirnen

Die hier vorgenommenen Charakterisierungen maschineller Lernstrategien bewegen sich auf der Betrachtungsebene konzeptioneller Ideen und basieren auf der Betrachtung der Ideen und Motivationen hinter der Entwicklung lernender Algorithmen. Die Entwicklung eines interdisziplinären Verständnisses der Lernstrategien und damit des maschinellen Lernens als Technikbereich hängt dementsprechend stark davon ab, dass die Grenzen der zugrunde liegenden Metaphern klar dargestellt werden. Dies stellt beim evolutionären Lernen ein relativ kleines Problem dar, weil die verkürzte Verwendung von Begriffen aus dem Kontext der Evolution ein häufig anzutreffendes Phänomen darstellt und die Begriffe automatisch mit einer gewissen Skepsis betrachtet werden. Künstliche neuronale Netze erfordern dieselbe Form von Skepsis und um dies zu begründen folgt ein kurzer Abriss eines wesentlichen Standpunktes innerhalb der Neuroanatomie bezüglich der Lernvorgänge im menschlichen Gehirn. Das Ziel hierbei ist, eine neutrale Betrachtung künstlicher neuronaler Netze zu ermöglichen und nicht eine Diskussion der Funktionsweise eines biologischen Gehirns vorzubereiten. Aus Sicht der Neuroanatomie ist einer der typischsten Fehler bei der Rede über das menschliche Gehirn, dass dessen Funktionsweise als mit der eines KNN vergleichbar verstanden wird (Teuchert-Noodt 2011). Diese These soll im Weiteren, aufbauend auf dem Standpunkt von Teuchert-Noodt, kurz begründet werden.

Die erste falsche Grundannahme liegt für Teuchert-Noodt darin, anzunehmen, dass Menschen als Kleinkinder in einer Art von **Nullzustand** ihr lebenslanges Lernen beginnen und dass Menschen Zusammenhänge kennenlernen, um sie dann durch eine Form von Wiederholung zu erlernen. In diesem Zusammenhang wird häufig die **Hebb'sche Lernregel** „what fires

³⁵ Parallelität bezieht sich hierbei auf die **konnektionistische** Idee der Darstellung eines Systems durch die massive Parallelisierung der Arbeitsschritte einfacher, vernetzter Einheiten und die damit verbundenen Möglichkeiten zu **verteilten Berechnungen** (Wikipedia Contributors 2012, Konnektionismus).

together, wires together“ genannt. Diese Aussage ist inhaltlich richtig, wird aber für das Verständnis von Lernprozessen deutlich überbewertet. Das menschliche Gehirn ist sehr stark vorstrukturiert. Selbst die Reihenfolge von möglichen Inhalten, die vom menschlichen Gehirn in den einzelnen Altersstufen der kindlichen Entwicklung erlernt werden können, ist stark vorgegeben.

Eine zweite falsche Intuition, die durch den Vergleich mit KNN vermittelt wird, ist diejenige, dass verstanden wird oder modellierbar ist, wie das menschliche Gehirn arbeitet. Dies ist schlicht noch nicht der Fall und die vorliegenden Erkenntnisse sind nicht ohne größere und noch ausstehende Anstrengungen in andere Wissenschaftsbereiche wie die Informatik zu übertragen. Im Gehirn liegen mit Neuronen und **Gliazellen** mindestens zwei unterschiedliche Formen von Nervenzellen vor, die eine zentrale Rolle spielen und in komplexen Abhängigkeiten zueinander stehen (Wikipedia Contributors 2012, Neuronales Netz). Die Funktionsweise dieser Gliazellen hat in der Arbeit mit KNN keine Entsprechung. Weiterhin sind die chemischen beziehungsweise hormonellen Abhängigkeiten innerhalb des Gehirns noch weitgehend unverstanden und insbesondere noch nicht mit Mitteln der Informatik modellierbar. Ein Grund für die Fokussierung auf Neuronen in der Diskussion der Funktionsweise des Gehirns könnte aus Sicht Teuchert-Noodts darin liegen, dass die Entdeckung von Nervenzellen in der Form von Neuronen zu genau prognostiziert wurde. Es lagen konkurrierende Theorien über die prinzipiellen Abläufe im Gehirn vor und eine dieser Theorien postulierte das Bestehen einer Zelle, die aufgebaut sein sollte, wie ein Neuron tatsächlich aufgebaut ist. Entsprechend war nach dem experimentellen Nachweis beziehungsweise der Entdeckung der Neuronen mittels verbesserter Technik die Überzeugung groß, dass die Theorie, die diese Zellen prognostiziert hatte, genau zutreffend sei. Dieser Erfolg hat die große Popularität der Neuronen mitbegründet und eine übergroße Emphase der Bedeutung dieser Art der Nervenzelle begünstigt, da Neuronen sofort intensiv untersucht werden konnten und inzwischen einen vergleichsweise gut dokumentierten Baustein des menschlichen Gehirns darstellen.

KNN sind ohne umfangreiche Weiterentwicklungen nicht dazu geeignet Gehirne beziehungsweise Gehirnprozesse zu simulieren. Zwar eignen sich KNN für den Einsatz in Kontexten, in denen ein Überfluss an Sensordaten vorliegt und bei denen wenig Vorwissen besteht, aber das Gehirn ist schlicht zu komplex um es ohne massive Vorstrukturierungen der einzusetzenden KNN funktionell anzunähern. Vereinfachende Theorien wie die Lokalisierung, die es erlauben würden, ein Gehirn mittels lokaler Modelle anzunähern, scheinen nicht einsetzbar zu sein. Zwar gibt es auch Erfolge im Versuch, Hirnprozesse mit Hilfe der Informatik zu verstehen, etwa bei der Diagnose von posttraumatischen Belastungsstörungen bei Soldaten (Hayes et al. 2011). Allerdings basieren diese Methoden, unabhängig davon, wie viel Potenzial ihnen überhaupt zugebilligt werden kann³⁶, nur sehr nachrangig auf maschinellem Lernen oder speziell auf KNN. Der Hauptvorteil des Einsatzes von KNN als MLA liegt – wie noch diskutiert wird – darin, dass KNN ohne umfangreiche Vorstrukturierung durch die Entwickler oder Nutzer eingesetzt werden können. Eine Simulation von Gehirnprozessen kann auf dieser Stärke nicht aufbauen.

Insgesamt sind künstliche Neuronen mathematische Modelle, die auf der Funktionsweise von natürlichen Neuronen basieren, aber so abgewandelt wurden, dass sie sich gut in der

³⁶ Eine Darstellung der engen Grenzen der Beobachtbarkeit von Hirnprozessen gibt Hasler (Hasler 2011, S. 39ff).

Informatik und dort im maschinellen Lernen einsetzen lassen. Zwar lassen sich einige Stärken und Schwächen von neuronalen Netzen auch bei künstlichen neuronalen Netzen beobachten, allerdings geben die neuronalen Netze nur einen Indikator dafür ab, in welchen Bereichen der KNN eine genauere Analyse gegebenenfalls einen Mehrwert ergeben würde.

Exemplarische Einsatzgebiete von KNN

Einer der bekanntesten und erfolgreichsten Einsätze künstlicher neuronaler Netze ist die Entwicklung von Algorithmen, die in der Lage sind, Backgammon zu spielen. Backgammon ist ein würfelbasiertes 2-Personen-Brettspiel, bei dem nach jedem Würfelwurf für den aktuellen Spieler eine begrenzte Anzahl von Zügen möglich ist. KNN können im Backgammon sehr erfolgreich als Lernstrategie eingesetzt werden.

[KNN] excel at strategic and positional judgment, using their knowledge to make fine distinctions between plays. They are less skilled in "technical" positions, such as bearing in against an anchor, which humans solve by calculation of the probabilities.

This is the opposite of the situation in many other games, in which computers calculate tactics well but fall short in strategic understanding. As a result there's been a lot of interest in applying temporal difference learning with neural nets to other games, and to mundane tasks too.

(Scott 2001)

Die Aussage von Scott zeigt eine für Algorithmen kontraintuitive Schwäche: sie sind menschlichen Spielern in berechenbaren Situationen unterlegen. Auch beim Backgammon treten solche Situationen auf, etwa im Endspiel, in dem die Spieler häufig keine Möglichkeit mehr haben die Spielsteine des Gegners zu bedrohen oder dessen Zugplanung zu beeinflussen. KNN sind schlechter als andere MLA für Kontexte geeignet, in denen Vorwissen wie die Wahrscheinlichkeiten von Würfelerggebnisse dem MLA mittels einer Vorstrukturierung vorgegeben werden können. KNN können in solchen Stellungen nicht analytisch die Würfelwahrscheinlichkeiten errechnen und einen optimalen Zug identifizieren. Eine in diesen Situationen von menschlichen Spielern umgesetzte, stellungsbezogene Spielweise beruht auf klaren Berechnungen oder zumindest mathematischen Abschätzungen der Nützlichkeit eines Zuges, basierend auf mathematischem Vorwissen. Dennoch sind KNN sehr erfolgreich im Backgammon eingesetzt worden, da dort die präzise Bewertung der aktuellen Spielposition – der **Stellung** – aufgrund fehlenden theoretischen Wissens vergleichsweise schwierig ist. Unabhängig von der Frage des fehlenden Vorwissens ist es bei der Nutzung eines KNN besonders gut möglich, das MLA im Rahmen des Autoadaptionsvorgangs gegen eine Kopie von sich selbst spielen zu lassen, ohne dass der Lernvorgang darunter leidet. Auf diese Weise können sehr viele simulierte Spiele in sehr kurzer Zeit durchgeführt werden und der Zufallseffekt des Würfelwurfes wirkt dabei einer Überanpassung entgegen. Der Autoadaptionsprozess benötigt daher im Prinzip keine Trainingsdaten und kann größtenteils automatisch ablaufen. Die Eingabedaten eines MLA sind in diesem Fall die Ergebnisse der Würfelwürfe und die Züge des anderen MLA. Insgesamt bewegt sich die Spielstärke der besten künstlichen neuronalen Netze im Backgammon auf Weltklassenniveau und einige der entsprechenden Programme sind kostenlos im Internet verfügbar. Die verfügbaren KNN sind dann bereits mit mehreren Millionen Partien trainiert worden und stellen schon den aus dem

Autoadaptionsprozess entstandenen Strukturvorschlag dar, wodurch sie auch auf rechen-schwachen Computern eingesetzt werden können³⁷.

Ein zweites Anwendungsbeispiel kommt aus der Bilderkennung (Russell et al. 2007, S. 914ff) und soll dazu dienen die Leistungsfähigkeit von KNN zu illustrieren. In dieser Anwendung wurde eine Datenbank aus 60.000 handschriftlichen Ziffernproben zugrunde gelegt und die Aufgabe bestand darin die Ziffern zu erkennen. KNN konnten ihre Fehlerrate im Laufe einiger Weiterentwicklungen von 1,6% über 0,9% auf schließlich 0,7% verbessern. Hier wurde sehr viel menschliches Vorwissen in den Autoadaptionsprozess eingebracht und die KNN wurden gezielt für den spezifischen Kontext vorstrukturiert. Die Fehlerrate eines Menschen bei der Ziffernerkennung liegt im genannten Beispiel geschätzt bei 0,2%, allerdings wurde für eine vergleichbare Datenbasis des United States Postal Service heuristisch eine Fehlerrate von 2,5% für den Menschen ermittelt. In jedem Fall geben die Größenordnungen der Fehlerraten ein sehr hilfreiches Gefühl für die Leistungsfähigkeit von KNN³⁸.

Funktionsbeschreibung künstlicher neuronaler Netze

Die Teilfunktionalität des Gehirns, die in KNN nachgebildet werden soll, wird, wie bereits angedeutet, durch die erwähnte Hebb'sche Lernregel beschrieben. Diese 1946 vom Psychologen Donald Hebb aufgestellte These beschäftigt sich mit Strukturen verbundener Neuronen, das heißt, mit neuronalen Netzen, und besagt sinngemäß, dass ein neuronales Netz lernt, indem bei gleichzeitiger Reizung zweier Neuronen die Stärke ihrer Verbindung vergrößert wird. Die relativ bekannte Kurzfassung dieser Lernregel ist die Formulierung „what fires together, wires together“. Der Hauptgrund, aus dem diese Lernregel als Grundlage einer Lernstrategie eingesetzt wird, findet sich in der folgenden Perspektive.

Viele der Modelle, die diskutiert wurden, beschäftigen sich mit der Frage, welche logische Struktur ein System besitzen muss, um eine Eigenschaft X darzustellen... Ein alternativer Weg, auf diese Frage zu schauen, ist folgender: Was für ein System kann die Eigenschaft X (im Sinne einer Evolution) hervorbringen? Ich glaube, wir können in einer Zahl von interessanten Fällen zeigen, dass die zweite Frage gelöst werden kann, ohne die Antwort zur ersten zu kennen. (Rosenblatt 1962)

(Görz et al. 2003, S. 11)

KNN können mit Hilfe eines Autoadaptionsprozesses, basierend auf künstlichen Neuronen und der Hebb'schen Lernregel, erwünschte Eigenschaften hervorbringen, ohne dass im Vorhinein bekannt ist, welche Struktur für die Realisierung dieser Eigenschaft notwendig ist. Daraus folgt sofort, dass KNN ihre Struktur in noch größerem Umfang als andere MLA autoadaptiv anpassen müssen. Die entstehenden Netze stellen Strukturvorschläge dar, denen keine geschickte Codierung des Kontextes und kein formales Vorwissen zugrunde liegen, sondern die als Gesamtstruktur entsprechend der Hebb'schen Lernregel systematisch auf

³⁷ Eine gute lesbare Einführung in die Konstruktion eines Backgammon-Programms auf Basis von KNN – inklusive einer nützlichen Visualisierung der Grenzen der Leistungssteigerung durch immer weitere Trainingspartien – findet sich bei Tsinteris (Tsinteris 2012).

³⁸ Einige gut verständliche, Animationen der Vorgehensweise beziehungsweise der Fähigkeiten von KNN im Hinblick auf Zahlen-erkennung finden sich bei LeCun (LeCun 2011).

Eingaben reagieren. KNN kommen damit der ursprünglichen in der Einleitung dargestellten Idee eines assoziativ lernenden Algorithmus sehr nahe.

Für den Einsatz von KNN als Lernstrategie ergibt sich daraus, dass zur Durchführung des Lernvorgangs – analog zur Verwendung von evolutionärem Lernen – nur sehr wenige Parameter oder gar analytische Hintergrundinformationen identifiziert oder quantifiziert werden müssen. Lediglich die Auswahl der relevanten Eingabegrößen ist notwendig, und im Gegensatz zu evolutionärem Lernen muss darüber hinaus kein Aufwand in die Erstellung einer Codierung investiert werden. Gerade weil dem Autoadaptionsprozess eines KNN keine interpretierbare Codierung zugrunde liegt, können aus dem Strukturvorschlag eines KNN nicht ohne Weiteres die Faktoren, die zu diesem Ergebnis geführt haben, abgelesen werden. Die Autoadaptation eines KNN resultiert in der Aneignung einer Fähigkeit und nicht in der Darstellung, wie diese Fähigkeit erlernt werden kann.

Die Komponenten künstlicher neuronaler Netze

Ein künstliches neuronales Netz setzt sich aus **Einheiten** genannten Knoten und **gerichteten Verbindungen** zwischen diesen Knoten zusammen. Die Einheiten sind nicht notwendigerweise mit allen anderen Einheiten verbunden, insbesondere liegt nicht immer eine Verbindung in beide Richtungen vor. Die Daten, die ein KNN erhält und weiterverarbeitet, werden **Signale** genannt. Dieser Begriffsbildung soll hier gefolgt werden, um die Intuition einer potenziellen Inhaltslosigkeit und der Konzeptlosigkeit eines Signals zu stärken. Ein KNN ist in der Lage Eingabesignale aufzunehmen, sie zwischen den Einheiten weiterzuleiten, dabei zu modifizieren und schließlich Ausgabesignale zu erzeugen. Die Stärke einer gerichteten Verbindung zwischen zwei Knoten wird deren **Gewicht** genannt. Der Autoadaptionsprozess eines KNN entspricht der Adaption der Gewichte der Verbindungen zwischen Einheiten, wobei sich prinzipiell jedes Gewicht in jedem Adaptionsschritt ändern kann und jede solche Änderung die Reaktionsmuster des gesamten KNN beeinflussen kann. KNN verzichten im Rahmen des Autoadaptionsprozesses auf die Manipulation interpretierbarer Symbole. Strukturvorschläge setzen sich nicht aus codierten Regeln zusammen, sondern aus einer Anordnung von Knoten, Verbindungen und Verbindungsgewichten. Die nachfolgende Abbildung zeigt eine exemplarische Visualisierung einer solchen Anordnung mit drei Eingabesignalen A, B und C, sowie einem Ausgabesignal D.

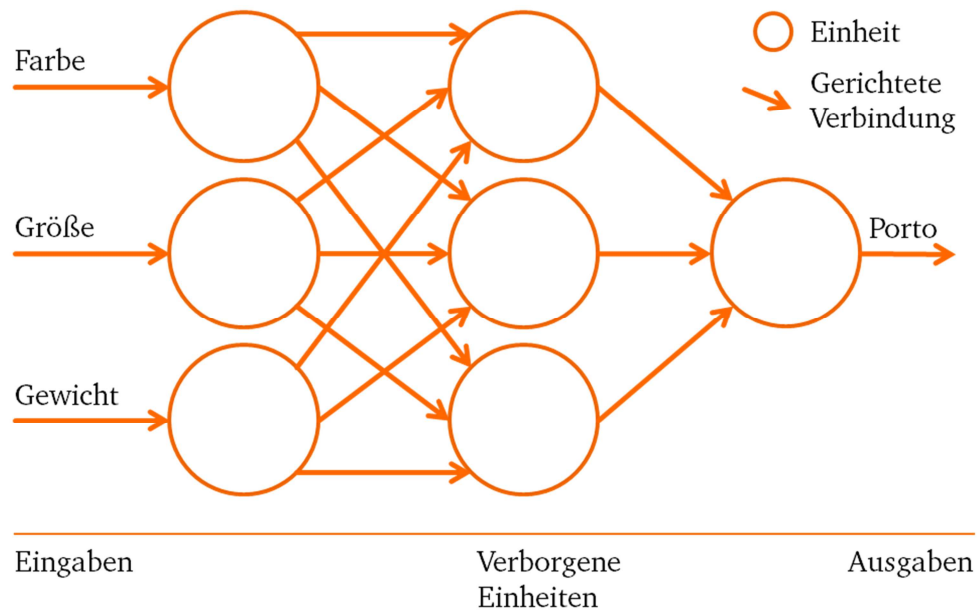


Abbildung 17: Vollständig verbundenes künstliches neuronales Netz

Als erstes soll die Rolle der Einheiten genauer betrachtet werden. Diese fungieren als **künstliche Neuronen**, das heißt sie orientieren sich in ihrer Funktionsweise an den Neuronen im menschlichen Gehirn.

Die „Schaltungstechnik“ von Neuronen kennt üblicherweise mehrere Eingangsverbindungen sowie eine Ausgangsverbindung. Wenn die Summe der Eingangsreize einen gewissen Schwellenwert überschreitet [...] „feuert“ das Neuron [...] das Ausgangssignal des Neurons.

(Wikipedia Contributors 2012, Neuronales Netz)

KNN konstruieren künstliche **Neuronen**, indem sie Einheiten einsetzen, die in der Lage sind aus einem oder mehreren Eingabesignalen ein oder mitunter auch mehrere Ausgabesignale zu erzeugen. Die Erzeugung eines Ausgabesignals in einer Einheit wird entsprechend als **Aktivierung** dieser Einheit bezeichnet. Aktivierungen können sowohl der Signalweiterleitung an eine andere Einheit als auch der Ausgabe an den Nutzer des KNN dienen. Einheiten, deren Aktivierungsfunktion ein Eingabesignal des Nutzers aufnimmt oder ein Ausgabesignal an den Nutzer abgibt, werden als **Eingabe-** respektive **Ausgabeeinheiten** bezeichnet. Die verbliebenen Einheiten werden unter dem Begriff **verborgene Einheiten** zusammengefasst. Im obigen Beispiel sind in der mittleren Spalte drei verborgene Einheiten zu sehen.

Die Systematik, nach der eine Einheit Aktivierungen vornimmt, wird als **Aktivierungsfunktion** bezeichnet. Eine Aktivierungsfunktion muss nur eine Anforderung erfüllen, sie muss die Entscheidung über die Aktivierung der zugeordneten Einheit nach einem systematischen Kriterium treffen, das die jeweiligen Eingabesignale berücksichtigt. Unterschiedliche Einheiten können individuelle Aktivierungsfunktionen aufweisen, und eine Anpassung dieser Funktionen im Rahmen des Autoadaptionsprozesses eines KNN ist zwar unüblich, kann aber durchaus vorgenommen werden. Eine einfache Aktivierungsfunktion besteht darin, die Stärke der Eingabesignale zu summieren und die jeweilige Einheit zu aktivieren, wenn diese Summe einen gewissen **Schwellenwert** überschreitet. Eine leichte

Weiterentwicklung dieser Aktivierungsfunktion besteht darin, die Eingabesignale nicht länger gleichberechtigt zu summieren, sondern jedes Signal in seiner Wichtigkeit für die Summe einzuschätzen und mit einem **Gewicht** zu belegen. Im der folgenden Visualisierung wurde einmal eine normale Summe gebildet und in zweiten Fall dem Signal B ein doppelt so großes Gewicht beigemessen wie dem Signal A.

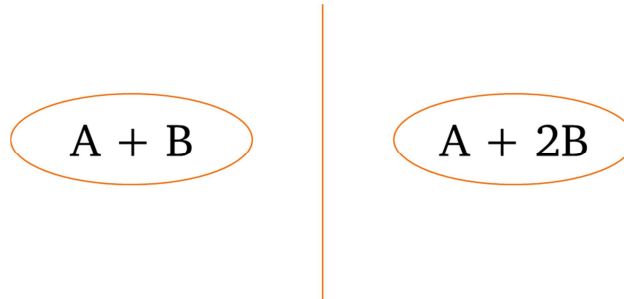


Abbildung 18: Zwei Möglichkeiten einer einfachen Signalgewichtung

Veränderungen in der Signalstärke von Signal B haben durch die Wahl dieses Gewichtes einen doppelt so großen Einfluss auf die Erreichung des Schwellenwertes wie Veränderungen von A. Jeder gerichteten Verbindung eines KNN, außer den Ausgabesignalen an die Nutzer, ist eine Gewichtung zugeordnet. Dieses **Verbindungsgewicht** stellt den Einfluss des transportierten Signals auf die angesteuerte Einheit dar. Die Anzahl der Einheiten und die Definition der Aktivierungsfunktionen, sowie die Anzahl und Orientierung der gerichteten Verbindungen eines KNN werden meist vor Beginn des Autoadaptionsprozesses fixiert. Die Adaptivität eines KNN liegt in diesem Fall ausschließlich in der Wahl der Verbindungsgewichte, das heißt die Verbindungsgewichte stellen die manipulierbaren Parameter des KNN dar. Der Raum aller möglichen Verbindungsgewichte stellt für die KNN den Suchraum des Autoadaptionsprozesses dar und wird als **Gewichtungsraum** bezeichnet. Die Verbindungsgewichte ändern sich während des Autoadaptionsprozesses mit der Betrachtung jedes Eingabedatums. Ein KNN kann mit zufälligen Gewichten initialisiert werden, optional kann jedoch auch analytisches Hintergrundwissen zu Abhängigkeiten zwischen den Attributen der Eingabedaten bei der Initialisierung der Gewichte berücksichtigt werden. Einerseits kann, wie im Beispiel der Ziffernerkennung, die Leistung eines KNN verbessert werden, wenn Hintergrundwissen eingesetzt wird, andererseits besteht der wesentliche Punkt gerade darin, dass diese Möglichkeit optional ist. Die Eingabedaten eines KNN können in nahezu jeder Art und Weise übergeben werden, es ist nicht erforderlich eine konsistente oder alle Teilaspekte erfassende Codierung zu erstellen. Die Eingabedaten dürfen und werden in der Praxis nicht direkt vergleichbar, fragmentarisch und mitunter sogar widersprüchlich sein. Das Lernen mittels künstlicher neuronaler Netze ist dementsprechend weitgehend unempfindlich auch gegenüber starkem Rauschen. Es genügt prinzipiell all diejenigen Aspekte der Trainingsdaten, die als potenziell relevant für den zu erstellenden Strukturvorschlag eingestuft werden, zu erfassen und dem KNN zu übergeben.

Der Autoadaptionsprozess bei KNN

Der Autoadaptionsprozess bei KNN durchläuft Zyklen, die als **Epochen** bezeichnet werden. Innerhalb einer Epoche besteht der Prozess aus den folgenden Schritten.

-
- A Die Eingabesignale werden von den Eingabeeinheiten aufgenommen. Die Eingabeeinheiten initiieren einen durch die Aktivierungsfunktionen und die Gewichte der beteiligten Einheiten geleiteten **Signalfluss**.
 - B Der Signalfluss setzt sich entlang der gerichteten Verbindungen durch das KNN fort.
 - C Die Ausgabeeinheiten geben die resultierenden Signale aus.
 - D Eine **Aktualisierungsregel** tritt in Kraft und nimmt eine Anpassung der Gewichte des KNN vor.

Die Aktivierungsregel kann sehr unterschiedlich ausfallen und sowohl überwachtes als auch unüberwachtes Lernen realisieren. Wenn ein KNN überwachtes Lernen realisieren soll, wird ihm ein Trainingsdatum übergeben und das Ausgabesignal, das das KNN im seinem aktuellen Zustand erzeugt, wird mit dem Wert verglichen, den das Trainingsdatum vorgibt. Die Abweichung wird als **Fehler** bezeichnet und die Gewichte des KNN werden so angepasst, dass der Fehler der Ausgabeeinheiten minimal oder zumindest kleiner als zuvor wird. Diese Vorgehensweise birgt die Herausforderung, dass Fehler verborgener Einheiten nicht direkt messbar sind. Zwar können verborgene Einheiten in den meisten KNN beobachtet werden, aber das Verborgene an ihnen ist ihre Funktion beziehungsweise ihre Relevanz für das Gesamtnetz. Die Trainingsdaten machen natürlich keine Vorgaben für den Zustand von verborgenen Einheiten. Ein Ansatz mit diesem Problem umzugehen, besteht darin, eine Einheit als für einen Teil der Fehler aller ihr nachfolgenden Einheiten anzusehen. Auf diesem Weg können Fehler von Ausgabeeinheiten auf die mit ihnen verbundenen verborgenen Einheiten übertragen werden, wodurch ein epochenabhängiger Sollwert für die entsprechende verborgene Einheit erstellt werden kann. Dieser Ansatz zur Analyse verbogener Einheiten wird als **Backpropagation** bezeichnet und soll im Weiteren visualisiert und detaillierter dargestellt werden. Ein Verständnis inwiefern die verborgenen Einheiten eines KNN verborgen sind, beziehungsweise analysiert werden können, verlangt zumindest grundlegende interdisziplinäre Kenntnisse bezüglich dieser Aktualisierungsregel. Zur besseren Übersicht wurde das KNN in der Visualisierung gegenüber der vorherigen Abbildung spiegelverkehrt dargestellt, da im ersten Schritt die Ausgabeeinheit betrachtet wird und dann entgegen des Signalflusses die übrigen Einheiten analysiert werden.

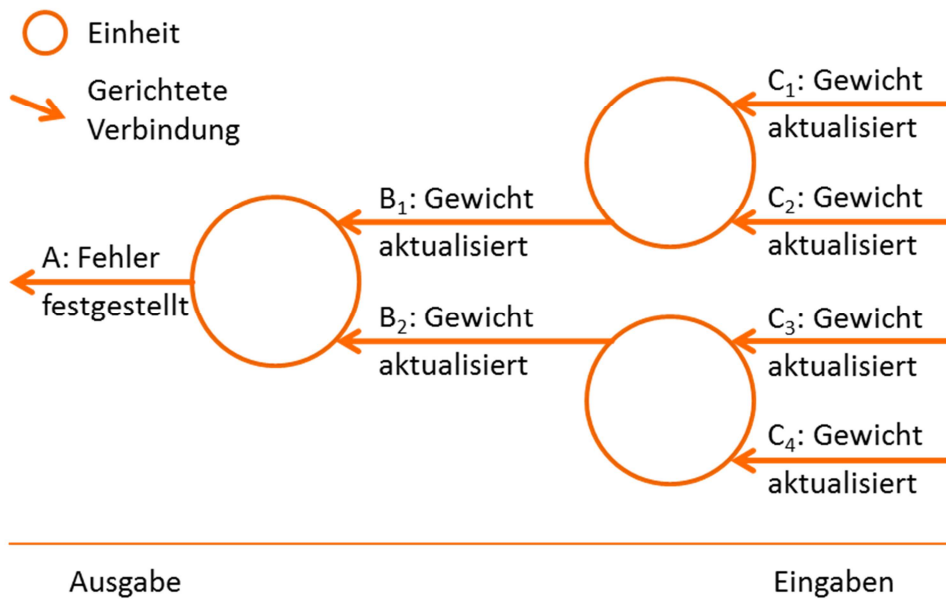


Abbildung 19: Backpropagation als Aktualisierungsregel überwachter Lernens

Die Aktualisierungsregel zur Adaption der Gewichte verborgener Einheiten basiert auf den folgenden Schritten.

- A** Beim Vergleich eines Ausgabesignals mit den Vorgaben eines Trainingsdatums wird ein Fehler festgestellt.
- B** Die Gewichte der eingehenden Verbindungen werden so abgeändert, dass der Fehler in der Ausgabeeinheit verkleinert wird.
- C** Die Verbindungen, deren Gewichte gerade verändert wurden, entstammen bestimmten Einheiten – im obigen Beispiel den zwei rechten Einheiten. Der Vergleich der gewünschten neuen Ausgabesignale mit den bisherigen Ausgabesignalen definiert einen Fehler. Die Gewichte der eingehenden Verbindungen werden wiederum so abgeändert, dass dieser Fehler verkleinert wird.
- D** Schritt **C** wird wiederholt, bis die Eingabeeinheiten erreicht werden.

Die genannten Schritte können und werden in der Praxis im Rahmen des Einsatzes einer Aktualisierungsregel sehr häufig durchlaufen. Das bedeutet, innerhalb einer Epoche wird die Aktualisierungsregel formal nur einmal angewendet, aber dennoch können sehr viele Gewichtsaktualisierungen vorgenommen werden. Ein Beispiel ist, dass der beschriebene Prozess zur Adaption der Gewichte solange durchlaufen wird, bis die zu korrigierenden Fehler eine festgesetzte Grenze unterschreiten.

Wenn im Zusammenhang mit KNN von unüberwachtem Lernen gesprochen wird, dann wird darunter verstanden, dass das KNN Signale klassifizieren kann, ohne dass die möglichen Klassen im Vorhinein bekannt sind. Ein erstes Beispiel hierfür sind **selbstorganisierende Merkmalskarten** oder kurz selbstorganisierte Karten. Hierbei wird jede Eingabeeinheit mit allen Nicht-Eingabeeinheiten des KNN verbunden, wie etwa im folgenden Beispiel.

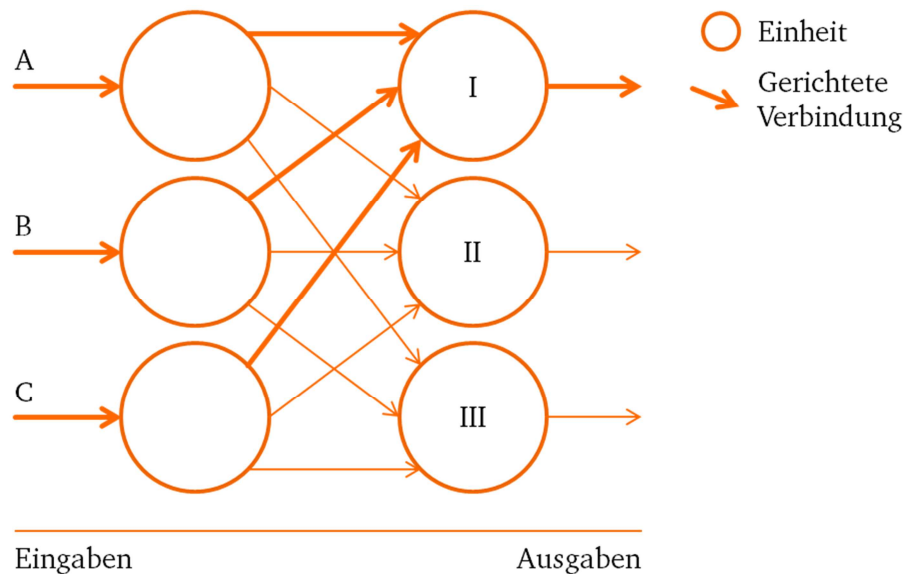


Abbildung 20: Selbstorganisierende Karte, die eine Eingabe der Kategorie I zuordnet

Das Konzept einer selbstorganisierenden Karte ist, dass für jedes Eingabesignal diejenigen Einheiten identifiziert werden, deren Verbindungsgewichte die größte Ähnlichkeit mit diesem Eingabesignal aufweisen. Im obigen Beispiel entspricht das aktuell betrachtete Trainingsdatum den Eingangssignalen A, B und C und die Verbindungsgewichte der Einheit I sollen am ehesten mit den Eingangssignalen übereinstimmen. Nachdem auf diese Weise eine Einheit identifiziert wurde, besteht die Aktualisierungsregel selbstorganisierender Karten darin, dass die Verbindungsgewichte der beschriebenen Einheit dem Eingabesignal noch weiter angenähert werden. Wenn immer wieder dasselbe Trainingsdatum eingelesen wird, wird irgendwann eine Einheit mit genau den zu den Eingangssignalen passenden Verbindungsgewichten vorliegen. Im obigen Beispiel wurde der Umgang mit drei Signalen dargestellt. Die Verbindungsgewichte einer Einheit können in diesem Beispiel als die Koordinaten eines Punktes im dreidimensionalen Raum interpretiert werden. Im Beispiel lagen drei Einheiten vor, die jeweils drei Verbindungsgewichte aufweisen. Das bedeutet, den Einheiten I, II und III können Punkte im Raum zugewiesen werden. Wenn beispielsweise angenommen wird, dass jede Ausgabeneinheit nur genau ein Verbindungsgewicht ungleich null besitzt, so liegen die drei Einheiten I, II und III in einem Koordinatensystem auf den Achsen³⁹.

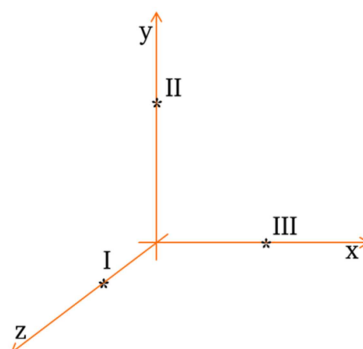


Abbildung 21: Verbindungsgewichte als Werte der Raumachsen Höhe, Breite und Tiefe

³⁹ Zum Beispiel hat die Einheit II eine Ausprägung von null in die Breite und in die Tiefe. Diese Einheit besitzt lediglich für die Höhe einen (in diesem Fall positiven) Wert.

Wenn jetzt ein Eingangssignal an das MLA übergeben wird, können die drei Komponenten A, B und C der Eingabe wiederum als Koordinaten im dreidimensionalen Raum interpretiert werden. Die Annäherung der dem Eingangssignal ähnlichsten Ausgabeeinheit könnte beispielsweise wie folgt aussehen.

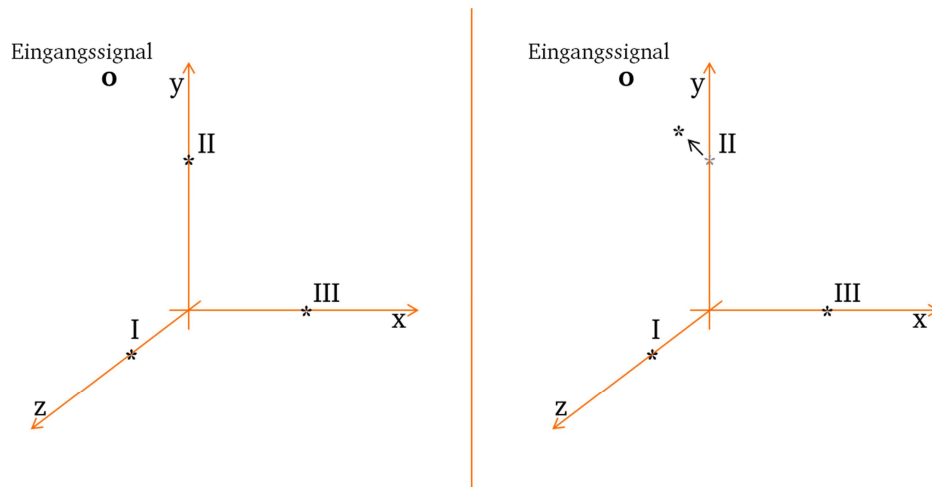


Abbildung 22: Adaption von Verbindungsgewichten mittels eines Eingangssignals

Die Visualisierung ist etwas irreführend, da sich in der Praxis bei einer selbstorganisierenden Karte die Gewichte sehr vieler Einheiten „in Richtung“ des Eingangssignals bewegen, die nächstgelegenen jedoch stärker als die weiter entfernten Einheiten. So bilden sich nach einer gewissen Zeit die Muster, die in den Rohdaten vorliegen, im KNN nach. Die Funktionsweise selbstorganisierter Karten liefert eine sehr gute Intuition, wie ein MLA in der Lage sein kann, nach Durchführung eines völlig ungesteuerten Autoadaptionprozesses Strukturen vorzuschlagen, die einen Bezug zu den Eingabedaten haben.

Lernen mittels KNN hat generell die Schwäche, dass Eingangssignale, die keine Ähnlichkeit zu den zuvor verarbeiteten Trainingsdaten aufweisen, ein unberechenbares Verhalten des KNN hervorrufen. Die **adaptive Resonanztheorie** ist ein Ansatz des unüberwachten Lernens, der dieses Problem zumindest für den Zeitraum des Lernvorgangs behebt. Allerdings reagieren auch die auf diese Weise erzeugten Strukturvorschläge – die Ergebnisse des Autoadaptionprozesses – nicht systematisch auf neuartige Eingangssignale.

Im Rahmen der adaptiven Resonanztheorie wird analog zu den selbstorganisierenden Karten bei der Aktualisierungsregel die Ähnlichkeit der Eingangssignale mit den Gewichten der Einheiten des KNN festgestellt und die ähnlichste Einheit dem Eingangssignal noch weiter angenähert. Allerdings werden nur die Gewichte der ähnlichsten Einheit adaptiert, die adaptive Resonanztheorie entspricht diesbezüglich genau dem obigen Beispiel. Über die Anpassung der ähnlichsten Einheit hinaus besitzt die adaptive Resonanztheorie, im Gegensatz zu den selbstorganisierenden Karten, einen zusätzlichen **Wachsamkeitsparameter**. Dieser Wachsamkeitsparameter entscheidet, ob die ähnlichste Einheit ausreichend große Übereinstimmungen aufweist oder ob mit dem Eingangssignal etwas komplett Neues vom MLA registriert wurde, für das im KNN noch keine Entsprechung besteht. Wenn die Ähnlichkeit als ausreichend groß betrachtet wird, wird das KNN als **in Resonanz** befindlich bezeichnet. In dem Fall, dass die Ähnlichkeit nicht ausreichend groß ist, wird eine zusätzliche Einheit erzeugt und die Gewichte der zusätzlichen Einheit werden entsprechend des als neu beurteilten Eingangssignals eingerichtet. Wenn Teile eines KNN sich zu einem Eingangssignal in

Resonanz befinden, können Einheiten eindeutig Klassen von Eingabesignalen zugeordnet werden. Das bedeutet, dass die Unterscheidung der Klassen von Eingabesignalen im entstandenen Strukturvorschlag eine strukturelle Entsprechung aufweist. Eine solche **lokale Repräsentation** hat zwei Stärken: zum einen kann ein solches KNN nach Abschluss des Autoadaptionsvorgangs aufgrund seiner hohen Parallelität sehr schnell ausgewertet werden und zum anderen besitzt es eine hohe Fehlertoleranz gegenüber Ausfällen einzelner Einheiten oder Verbindungen. Eine direkt damit zusammenhängende Schwäche besteht jedoch in dem vergleichsweise großen Zeitaufwand, der nötig ist um die Trainingsdaten zu lernen. Wenn aufgrund von Vorwissen die Anzahl der zu identifizierenden Eingabesignale bekannt ist oder abgeschätzt werden kann, wird in der Praxis häufig ein KNN mit einem entsprechend großen Reservoir an speziell ausgezeichneten Einheiten erzeugt, die nur adaptiert werden dürfen, wenn ein neues Eingabesignal erkannt wurde. Diese Vorstrukturierung hat den Vorteil, dass keine neuen Einheiten erzeugt werden müssen und dennoch eine lokale Repräsentation möglich ist. Die Idee hinter dieser Maßnahme ist auch über die adaptive Resonanztheorie hinweg von Bedeutung, da sie erklärt, wie prinzipiell im Vorfeld Einfluss auf den Aufbau eines KNN genommen werden kann. Darüber hinaus deutet sich hier an, dass und wie ein KNN ohne Vorstrukturierungen oder Steuerung in Reaktion auf die Eingabesignale systematisch wachsen und schrumpfen kann.

Darstellungskraft von KNN

Die Motivation zur Erstellung eines KNN lag unter anderem in der Idee, mittels eines Autoadaptionsprozesses Strukturvorschläge zu erstellen, die gewisse Eigenschaften aufweisen oder Funktionen erfüllen, ohne dass im Vorfeld klar sein muss, welche Voraussetzungen die Strukturvorschläge erfüllen müssen, um ebendies leisten zu können. Zwar verringert eine solche Vorgehensweise das notwendige Vorwissen, sie entbindet jedoch nicht von der Betrachtung, welche Eigenschaften und Funktionen ein KNN prinzipiell ausbilden kann. Kenntnisse über die Potenziale und Grenzen des Autoadaptionsprozesses sind notwendig um KNN konzeptionieren zu können, insbesondere um die Aktivierungsfunktionen der Einheiten festzulegen. Ein Beispiel für die Grenzen des Autoadaptionsprozesses bilden Aktivierungsfunktionen, die – wie im obigen Beispiel – eine Summe mit gewichteten Summanden bilden und prüfen, ob ein Schwellenwert überschritten wurde. Solche Aktivierungsfunktionen erlauben dem entstehenden KNN unabhängig vom übrigen Autoadaptionsprozess nur die Darstellung von mathematisch äußerst einfachen Funktionen und Eigenschaften. Zwar können komplexere Zusammenhänge näherungsweise durch einfachere Funktionen und Eigenschaften beschrieben werden, aber die Frage, welche Funktionen von welchen KNN prinzipiell darstellbar sind, ist dennoch von großer Bedeutung. Zur Beantwortung dieser Frage sollen zunächst zwei Klassen von KNN unterschieden werden: KNN, bei denen keine Rückkopplung erlaubt ist, die **azyklischen KNN**, und diejenigen Netze, bei denen Rückkopplungen zugelassen sind, die **rekurrenten KNN**.

Zuerst soll eine Darstellung der deutlich weniger komplexen azyklischen KNN vorgenommen werden. Die Ausgaben dieser KNN hängen nur von den Eingabesignalen und dem Zustand der Gewichte ab. Sie besitzen neben den Verbindungsgewichten keine veränderlichen Parameter und abgesehen von der zwischenzeitlichen Adaption der Gewichte gemäß der Aktualisierungsregel reagieren sie auf die gleichen Eingabesignale immer auf die gleiche

Weise. Jedes azyklische KNN kann formal in **Schichten** angeordnet dargestellt werden⁴⁰, wobei Einheiten jeder Schicht Signale nur aus der jeweils vorhergehenden Schicht empfangen und Signale nur in die folgende Schicht senden. Eingabeeinheiten werden dabei nicht als eigene Schicht betrachtet, da sie Eingabesignale ohne Gewichtung aufnehmen, diese aufspalten oder vervielfältigen und an andere Einheiten weiterleiten. Schichten werden deshalb in **verborgene Schichten** und die **Ausgabeschicht** unterteilt.

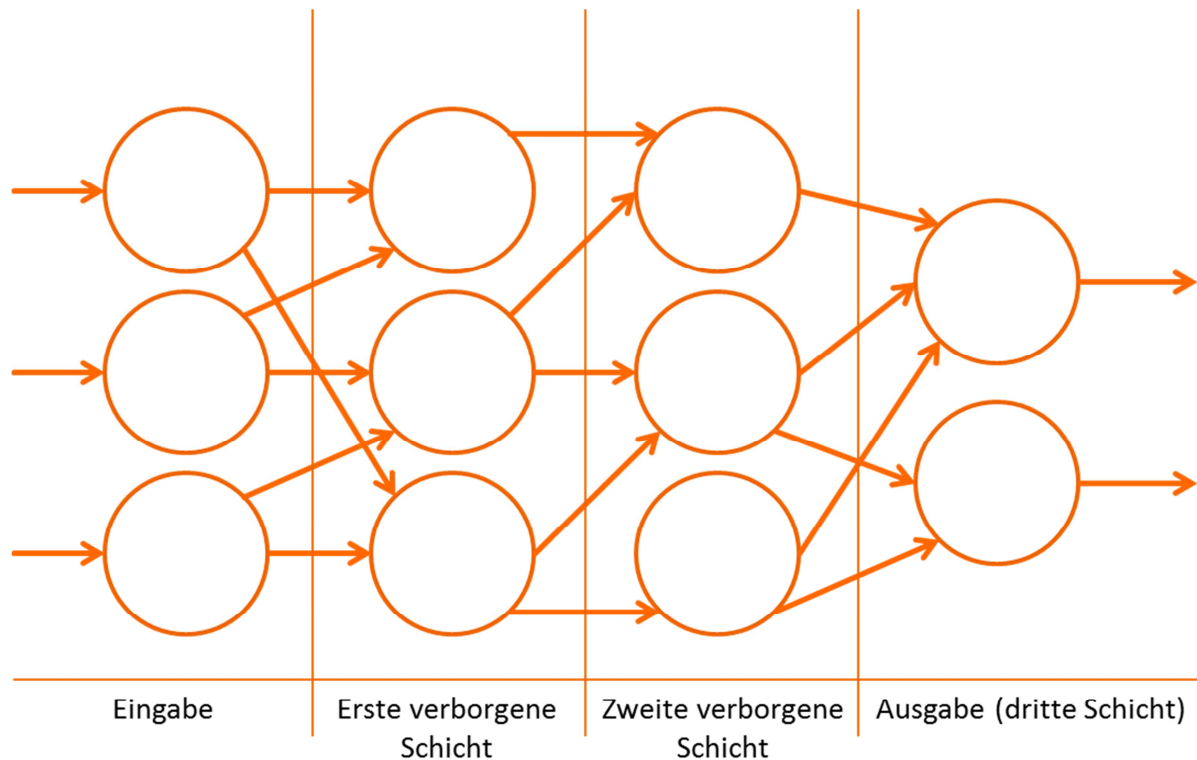


Abbildung 23: Übersicht der Schichten eines dreischichtigen KNN

Die Darstellungskraft von KNN wird beurteilt, indem betrachtet wird, welche mathematischen Funktionen und Operationen die Netze darstellen können, da die Umwandlung von Eingabesignalen in Ausgabesignale formal unabhängig von der konkreten Codierung ein mathematischer Vorgang ist.

Das einfachste azyklische Netz wird als **einlagiges Perzeptron** bezeichnet und besitzt nur eine Schicht, die dadurch gleichzeitig die Ausgabeschicht ist. Ein einlagiges Perzeptron kann den Raum der Eingabedaten **linear** in zwei Teile teilen, wenn als Aktivierungsfunktion, wie oben beschrieben, eine gewichtete Summe verwendet wird.

⁴⁰ Gegebenenfalls müssen dabei für Einheiten, die mit einer entfernten Schicht kommunizieren in den dazwischenliegenden Schichten Einheiten mit fixen Gewichten eingefügt werden, die lediglich das Signal weiterleiten.

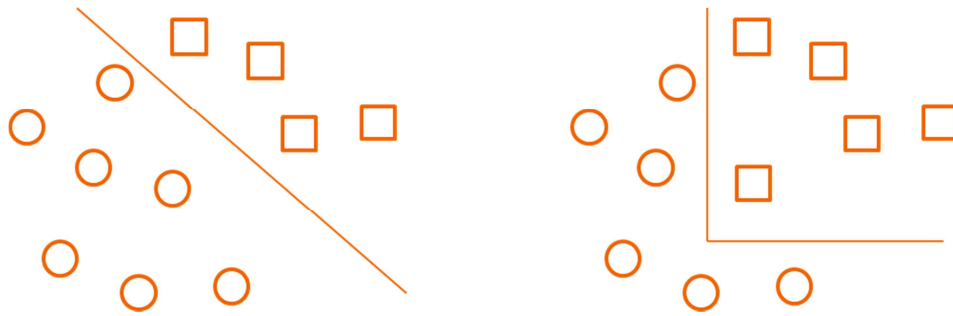


Abbildung 24: Lineare Trennung (links) und nicht-lineare Trennung (rechts)

Ein **zweilagiges Perzeptron**, das heißt ein KNN mit einer verborgenen Schicht, kann die meisten **mathematischen Funktionen** beliebig genau beschreiben⁴¹, wenn etwas komplexere Aktivierungsfunktionen eingesetzt werden. Auch beim Einsatz einer gewichteten Summe als Aktivierungsfunktion kann ein zweilagiges Perzeptron jedoch schon rein mittels der Adaption ihrer Gewichte die grundlegenden **logischen Funktionen** UND, ODER und NICHT abbilden. Nachfolgend wird zur Anschauung eine Einheit visualisiert, die ein UND abbildet. Die zwei Eingabesignale können dabei das Vorliegen zweier Eigenschaften codieren, wodurch die Einheit überprüft, ob die Eigenschaften gleichzeitig auftreten. Die zwei Eingabesignale werden so codiert, dass sie jeweils entweder eine 0 oder eine 1 übermitteln und die Entscheidung über die Aktivierung der Einheit wird getroffen, indem eine Summe der Eingabesignale mit dem Schwellenwert 1,5 verglichen wird.

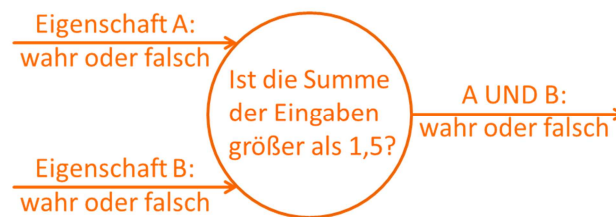


Abbildung 25: Realisierung der logischen Funktion UND mittels Verbindungsgewichten

Die beiden Gewichte der Summe sind dabei auf 1 festgelegt. Eine Senkung des Schwellenwertes auf $\frac{1}{2}$ oder eine Erhöhung der beiden Gewichte auf 2 kann diese Einheit zu einem ODER⁴² machen, da dann in beiden Fällen bereits ein einzelnes Eingangssignal ausreichen würde, um den Schwellenwert zu überschreiten und die Aktivierung der Einheit auszulösen. Die Ausbildung eines interdisziplinären Verständnisses von MLA setzt nicht voraus, KNN auf der Ebene der Wahl von Schwellenwerten zu verstehen. Das genannte Beispiel ist dennoch von einiger Bedeutung, weil es einen Eindruck vermittelt, wie im Rahmen eines einfachen Teilschrittes eines Autoadaptionsprozesses ohne einen Steuerungseingriff aus einem UND ein ODER werden kann. Dies ist ein verhältnismäßig konkretes Beispiel dafür, was bei MLA unter Selbstorganisation verstanden werden kann.

⁴¹ Mathematische Funktionen beschreiben jegliche Formen von eindeutigen Zuordnungen zwischen einem Eingabe- und genau einem Ausgabewert. Bestimmte, besonders einfache (präzise: beschränkte und stetige) mathematische Funktionen können von KNN aus zwei Schichten beliebig genau beschrieben werden, wobei die verborgene Schicht aus **sigmoiden Einheiten** zusammengesetzt sein muss. Der Grund ist, dass einfache (präzise: stetige) Funktionen als stückweise linear betrachtet werden können und die Zerlegung der anzunähernden Funktion je nach Forderung an die Genauigkeit immer kleiner gewählt werden kann.

⁴² Ein mathematisches ODER entspricht einem einschließenden ODER, das heißt einem „Und-oder“ und gerade keinem Entweder-oder.

Ein drei- beziehungsweise mehrlagiges Perzeptron schließlich kann – bei Verwendung der angedeuteten, etwas komplexeren Aktivierungsfunktion – bereits alle mathematischen Funktionen von praktischer Relevanz beliebig genau annähern⁴³.

Die Fähigkeit mehrschichtiger Netze, automatisch mathematische Funktionen mittels verborgener Schichten erstellen zu können, ermöglicht einen beträchtlichen Grad von Flexibilität bei der Suche nach Strukturvorschlägen, da diese nicht im Vorfeld vom Nutzer vorgegeben werden müssen. Entsprechend können KNN Strukturen vorschlagen, die dem Nutzer völlig unbekannt sind – und gegebenenfalls auch nach Aufbau beziehungsweise Adaption der verborgenen Schichten unbekannt bleiben, da verborgene Schichten nicht ohne Weiteres verständlich oder gar selbsterklärend sind. Hinzu kommt der systematische abduktive Bias azyklischer KNN, der in etwa darin besteht, dass angenommen wird, dass eine gleichförmige Annäherung an die gesuchte Struktur möglich ist. Meist wird diese Annahme jedoch vom Nutzer des MLA geteilt, wodurch der Bias an Bedeutung verliert. Mitunter können darüber hinaus bis zu einem gewissen Grad mögliche „Bedeutungen“ für die verborgenen Schichten gefunden werden, etwa wenn eine lokale Repräsentation vorliegt oder eine Analyse zeigt, dass eine spezielle verborgene Einheit einer Bilderkennung die Eigenschaft „links ist es sehr hell“ codiert.

Während allgemeine Aussagen der oben dargestellten Form über die Darstellungskraft von azyklischen KNN möglich sind, gilt dies nur sehr eingeschränkt für Aussagen bezüglich konkreter Funktionen. Im Einzelfall ist es sehr kompliziert, analytisch für eine Funktion oder eine Funktionenklasse zu bestimmen, wie viele verborgene Einheiten und Verbindungen genau benötigt werden, um die Funktion annähern oder abbilden zu können. Diese Schwäche spielt jedoch keine Rolle, wenn Suchräume betrachtet werden, über die im Vorfeld sehr wenig bekannt ist. In solchen Fällen ist unabhängig von der eingesetzten Lernstrategie unbekannt, wie das MLA genau vorstrukturiert werden muss. Varianten evolutionären Lernens, die in der Lage sind ihre Strukturvorschläge sehr stark zu verändern und KNN mit ihrer sehr großen Darstellungskraft eignen sich besonders gut für Einsätze in solchen Kontexten.

Alternativ zu azyklischen KNN kann die Erzeugung von Schleifen beziehungsweise Rückkopplungen im Rahmen des Autoadaptionprozesses auch erlaubt sein. Ein solches **rekurrentes KNN** gibt Teile seiner Ausgabesignale als Eingabesignale an sich selbst weiter. Die Gewichte dieser Art von KNN bilden ein **dynamisches System**, dessen unterschiedliche Reaktionsweisen sich meist mittels einer gewissen Zahl systematisch unterschiedlicher Zustände beschreiben lassen. Zu jedem Zeitpunkt kann sich das KNN entweder in einem chaotischen Zustand, einem stabilen Zustand oder einem schwingenden Zustand befinden. Ein chaotischer Zustand beschreibt eine zufällig erscheinende Reaktionsweise, ein stabiler Zustand eine Reaktionsweise analog zu einem azyklischen KNN und ein schwingender Zustand entspricht einem KNN, das sich wie ein Pendel zwischen mindestens zwei unterschiedlichen Reaktionsweisen hin und her bewegt. Die Antwort des KNN ist entsprechend abhängig von seinem derzeitigen Zustand, der wiederum von früheren Eingaben abhängig ist. Die Unterscheidung zwischen der Änderung eines Zustandes und einer Autoadaption kann verglichen werden mit der Nutzung eines Lichtschalters. Ein Lichtschalter hat meist zwei Zustände und abhängig von diesen Zuständen reagiert er auf eine Betätigung

⁴³ Insbesondere können unstetige Funktionen angenähert werden. Dies wird plausibel, wenn klar ist, dass zwei Schichten sigmoider Einheiten bereits alle stetigen Funktionen abbilden können und dass unstetige Funktionen sich durch Linearkombinationen von lokal definierten stetigen Funktionen darstellen lassen.

mit dem Ein- oder dem Ausschalten des Lichtes. Diese Zustände können jedoch nicht sinnvoll als das Ergebnis eines systematischen Autoadaptionsprozesses beschrieben werden.

Die Veränderung eines rekurrenten KNN unterscheidet sich auch insofern von den Adaptionen anderer KNN oder auf anderen Lernstrategien basierenden MLA, als der Autoadaptionsprozess zunächst ein Eingabedatum registriert und meist in einem gesonderten Schritt eine Aktualisierung oder Adaption vornimmt. Rekurrente KNN verändern sich mitunter schon bei der Registrierung von Eingabesignalen. Im Falle des Vorliegens eines schwingenden oder chaotischen Zustandes kann diese Veränderung der Antwort des dynamischen Systems – des Strukturvorschlags – sogar beliebig lange andauern. Denkbar wäre etwa ein KNN, das in Reaktion auf ein Eingabesignal seinen Zustand verändert, ein Ausgabesignal erzeugt und dieses wieder als Eingabesignal aufnimmt. Die Abhängigkeit von vergangenen Eingaben tritt dementsprechend nicht nur als systematische oder zumindest bewertete Anpassung im Rahmen des Autoadaptionsprozesses auf.

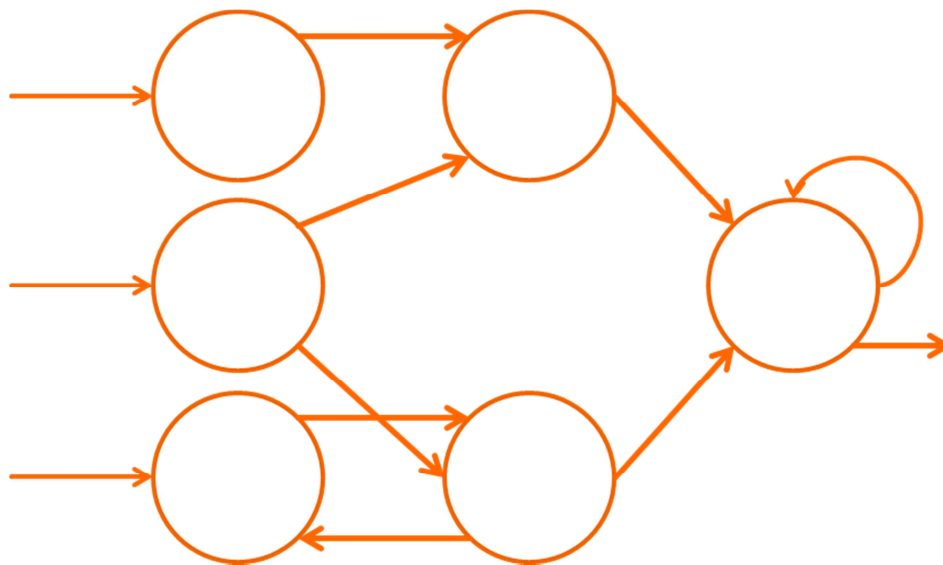


Abbildung 26: Rekursive Verbindung in einem rekurrenten KNN

Rekursive Netze sind als dynamische Systeme schwieriger mittels einzelner Trainingsdaten zu formen als azyklische Netze und der Autoadaptionsprozess verläuft weniger systematisch, allerdings können manche Abhängigkeiten innerhalb der betrachteten Rohdaten von rekursiven Netzen besser modelliert werden. Zur Erzeugung eines systematischeren Verhaltens können rekursive Netze durch azyklische Netze dargestellt werden, indem das betrachtete rekursive Netz vervielfältigt wird und rekursive Verbindungen jeweils in die nächste Kopie des Netzes eingehen. Der mit dieser Maßnahme verbundene Aufwand führt dazu, dass versucht wird, das Problem zu umgehen und nicht zu lösen. Ein Beispiel für solch einen Umweg besteht darin, einen evolutionären Algorithmus einzusetzen, der KNN codiert. Das Problem hierbei ist, dass in diesem Fall zwar formal KNN verwendet werden, allerdings nicht als Grundlage für eine Lernstrategie, sondern lediglich als Darstellungsform eines rekurrenten Zusammenhangs. Die Kenntnis weiterer Details zu rekurrenten KNN ist für ein interdisziplinäres Verständnis nicht vonnöten. Allerdings ist es hilfreich zu wissen, dass auch die hier nur angedeuteten rekurrenten KNN noch weit von der Komplexität biologischer neuronaler Netze entfernt sind. Biologische neuronale Netze haben beispielsweise die zusätzliche Anforderung, dass Neuronen ein Membranpotenzial besitzen, das ausreichend groß sein muss, wenn das entsprechende Neuron in der Lage sein soll zu feuern.

Entsprechend spielt der genaue Zeitpunkt des Feuerns eines Neurons in biologischen neuronalen Netzen eine bedeutende Rolle. Diese und andere Eigenschaften lassen sich abstrakt als Erweiterung in die Idee von dynamischen Systemen aufnehmen und werden im Rahmen von komplexeren KNN modelliert.

Stützen der Netzstruktur gegen Überanpassung

In der bisherigen Betrachtung wurde der Lernvorgang eines KNN meist mit der Modifikation seiner Gewichte identifiziert. Dies lässt sich erweitern, indem das KNN den eigenen Aufbau als Netz ebenfalls adaptiert. Die Neuschaffung oder die Entfernung von Verbindungen zwischen Knoten und von Einheiten kann in den Autoadaptionsprozess aufgenommen werden anstatt beides vor Beginn des Prozesses zu fixieren. Diese Vorgehensweise wurde bei der Darstellung der adaptiven Resonanztheorie bereits angedeutet. Dort bestand die Möglichkeit, lokale Repräsentationen zu realisieren, indem die entsprechenden Einheiten neu erstellt werden oder im Vorfeld als solche ausgewählt werden (Fahlman 1991). Wenn ein KNN im Rahmen des Autoadaptionsprozesses erweitert werden soll, kann eine verborgene Einheit ergänzt und deren Gewichte heuristisch so eingestellt werden, dass der Fehler des erweiterten Netzes minimal ist. Anschließend werden die Gewichte der neu ergänzten Einheit fixiert, während die Gewichte des ursprünglichen Netzes adaptiert werden. Dies wird mehrfach wiederholt. Eine solche Erweiterung von KNN ermöglicht eine sehr schnelle Adaption an neuartige Signale. Allerdings liegt eine naheliegende Gefahr in der übermäßigen Ergänzung verborgener Einheiten und damit einer Überanpassung an die Trainingsdaten. Als Gegenmaßnahme lassen sich die Anzahl und Art der Verknüpfungen zwischen den Einheiten und die Anzahl der Einheiten auch wieder reduzieren. Die Vorgehensweise ähnelt dabei dem Stützen von Entscheidungsbäumen. Ein vollständig vernetztes KNN wird erstellt und es werden Verbindungen zwischen Einheiten oder ganze Areale identifiziert, deren Relevanz für das Gesamtnetz fraglich ist. Ein Grund kann sein, dass sich die Gewichte im entsprechenden Gebiet während des gesamten Autoadaptionsprozesses kaum verändert haben. Die Auswirkung des Teilnetzes auf die Performanz des Gesamtnetzes kann überprüft werden, indem ein zweites KNN erstellt wird, dem die identifizierten Verbindungen oder Areale fehlen. Dieses zweite KNN kann anschließend mit dem ursprünglichen KNN auf Performanz oder bezüglich anderer Kriterien wie der Antwortgeschwindigkeit verglichen werden⁴⁴. Ein Hauptziel, das mit der Stützung von KNN verfolgt wird, ist die Vermeidung oder Reduzierung von Überanpassungen. Ein großes KNN kann genau wie ein großer Entscheidungsbaum alle Trainingsinstanzen reproduzieren, indem es in den verborgenen Schichten eine Art Nachschlagetabelle anlegt. Kurz gesagt verringert sich die Tendenz zur Überanpassung mit sinkender Anzahl von verborgenen Einheiten. Eine Möglichkeit sehr große KNN zu vermeiden, anstatt sie zu reduzieren, besteht darin, aus mehreren unabhängigen kleinen Netzen große Netze zusammen zu setzen. Jedes der betrachteten kleinen Netze muss in diesem Fall bereits isoliert möglichst viele Trainingsdaten erklären können. Anschließend werden solange kleine Netze, die disjunkte Mengen von Trainingsdaten erklären, zusammengeschaltet, bis alle Trainingsdaten abgedeckt sind.

Insgesamt kann der Autoadaptionsprozess eines KNN, wie schon derjenige bei evolutionärem Lernen, sehr weitreichende Veränderungen des Strukturvorschlages bewirken. Im Gegensatz zu evolutionärem Lernen wird bei KNN allerdings keinerlei Wert auf eine Codierung gelegt

⁴⁴ Dieser Vergleich unterschiedlich aufgebauter KNN kann so intensiv betrieben werden, dass für einen bestimmten Kontext heuristisch eine optimale Anzahl von Schichten und der Knoten pro Schicht bestimmt werden können.

und eine prinzipielle Unverständlichkeit der Prozesse wird hingenommen, um eine maximale Darstellungskraft zu gewinnen. Ein interdisziplinäres Verständnis der Funktionsweise von künstlichen neuronalen Netzen hängt stark von der Vermeidung von Denkfehlern ab, daher wurde für die Darstellung der Motivation und der Funktionsbeschreibung in dieser kurzen Diskussion der Lernstrategie ein vergleichsweise großer Aufwand betrieben. KNN sind jedoch auch für die meisten technikphilosophischen Diskussionen eines MLA oder eines „selbstorganisierten Algorithmus“ von Bedeutung und können häufig als Beleg oder Gegenbeispiel für eine These eingesetzt werden.

Zusammenfassung des Zweckes von KNN

Zusammengefasst lässt sich sagen, dass künstliche neuronale Netze einen allgemeinen, praktischen Ansatz darstellen, um auf Basis von Messwerten Strukturvorschläge zu erstellen, die Funktionen abbilden, die mit einer großen Zahl von Eingangsgrößen agieren. Zum Einsatz kommende Algorithmen wie die **Backpropagation** benutzen iterative Methoden, um die Parameter von künstlichen neuronalen Netzen so einzustellen, dass diese möglichst performant auf einer Menge von Trainingsdaten sind, die in Form von Eingabe-Ausgabe-Paaren vorliegen. Autoadaptionsprozesse auf Basis von KNN sind unempfindlich gegenüber Rauschen und werden erfolgreich für die Bearbeitung einer Vielzahl von Problemen eingesetzt⁴⁵.

Trainingsdaten, die an ein KNN übergeben werden und die eine Modifikation der Verbindungsgewichte bewirken sollen, sind typischerweise durch eine große Anzahl von mit Zahlenwerten versehenen Attributen codiert. Messwerte zu diesen Attributen können dem KNN in nahezu jeder Art und Weise übergeben werden. Es ist nicht erforderlich, eine konsistente oder alle Teilaspekte der Problemumgebung erfassende Codierung zu erstellen, die Codierung darf und wird typischerweise fragmentarischen Charakter haben. Die Attribute müssen weder unabhängig noch korreliert sein, sie müssen nicht einmal vergleichbar oder widerspruchsfrei sein. Das bedeutet, eine Messung muss lediglich all diejenigen Aspekte der Trainingsinstanzen, die als potenziell relevant für das unbekannte und zu lernende Konzept eingestuft werden, als Zahlenwerte erfassen und dem KNN anschließend in beliebiger – wenn auch über alle Instanzen konstanter – Reihenfolge übergeben. Diese Robustheit gegenüber der Übergabereihenfolge von Messwerten veranschaulicht die hohe Fehlertoleranz von KNN gegenüber verrauschten Eingabewerten, aufgrund derer KNN sich besonders gut zur Darstellung und Verarbeitung von Unschärfe sowie Rauschen eignen. Die Fähigkeit künstlicher neuronaler Netze, auch in Kontexten einsetzbar zu sein, die dem Nutzer fremd oder völlig unbekannt sind, wird im zweiten Hauptteil eine zentrale Rolle spielen.

2.3.5. Instanzenbasiertes Lernen

Motivation

Die zentrale Motivation des **instanzenbasierten** oder **deskriptiven Lernens** besteht darin, die Trainingsdaten beziehungsweise Instanzen den Strukturvorschlag direkt und möglichst stark beeinflussen zu lassen. Die Trainingsdaten sollen möglichst unmittelbaren Einfluss auf die Klassifizierung neuer Eingabedaten haben. Insbesondere sollen die Trainingsdaten gerade *nicht* nur dazu genutzt werden, einen Strukturvorschlag in Form eines Baumes oder eines

⁴⁵ Eine sehr gut zugängliche und frei verfügbare Einführung in die technischen Details und die Möglichkeiten zur Implementierung von KNN findet sich bei Kriesel (Kriesel 2007).

Netzes zu erstellen. Nach Abschluss der Erstellung eines Entscheidungsbaumes besitzen die Trainingsdaten etwa allenfalls noch eine implizite Bedeutung. Die wichtigste Vorannahme des instanzbasierten Lernens muss dementsprechend sein, dass neue Ereignisse sehr wahrscheinlich den bereits bekannten Ereignissen ähneln. Dies kann gesteigert werden bis hin zu der Annahme, dass grundsätzlich keine Überraschungen auftreten und alle neu registrierten Eingabedaten immer bereits zuvor registrierten Eingabedaten ähneln. Diese Annahme ist verwandt zum Bias bei KNN, dort war die implizite Grundannahme, dass die Strukturunterschiede innerhalb der Rohdaten nicht extrem groß sind. Beide Annahmen stellen Schwächen der jeweiligen Lernstrategien dar, mit denen jedoch genau gegensätzlich umgegangen wird. Während KNN die Annahme als impliziten systematischen Fehler ignorieren, wird sie beim instanzbasierten Lernen zum zentralen Merkmal der Lernstrategie.

Die Idee der Klassifizierung von Eingabedaten mittels eines direkten Bezugs zu den Trainingsdaten war ebenfalls im Rahmen der Darstellung KNN bereits aufgetaucht. Dort wurden beim Konzept der adaptiven Resonanztheorie lokale Repräsentationen von Eingabedaten erzeugt. Die adaptive Resonanztheorie hatte dabei in erster Linie versucht, für jedes neu hinzugekommene Eingabedatum einen Abgleich mit den bisherigen Klassen oder klassifizierten Instanzen vorzunehmen, um die gebildeten Klassen den betrachteten Instanzen anzunähern. Diese Vorgehensweise steht beim instanzbasierten Lernen im Fokus. Die Diskussion instanzbasierter Lernstrategien kann dementsprechend genutzt werden, um andere Lernstrategien, die diesen Anspruch nicht explizit formulieren, anders und mitunter besser zu verstehen.

Ein Problem der Kategorisierung dieser Klasse von MLA und des nachfolgenden **statistischen Lernens** liegt darin, dass sich der Fokus von der Beschreibung einer Strategie zur Erstellung eines Autoadaptionsprozesses entfernt. Stattdessen steht tendenziell eine Strategie zur Erstellung von Strategien im Vordergrund. Dieses Problem ist jedoch nicht sehr schwerwiegend, da auch evolutionäres Lernen eine Gruppe von Ansätzen umfasste, die stark verwandt waren. Die Methoden, mit denen instanzbasiertes Lernen realisiert wird, werden im Weiteren allerdings nicht so detailliert betrachtet, wie die genetischen Algorithmen, die genetische Programmierung und die Evolutionsstrategien als Varianten evolutionären Lernens. Der Grund hierfür ist, dass die eingesetzten Methoden aus mathematischer Sicht relativ geradlinig sind und in erster Linie versuchen die konzeptionellen Vorgaben des instanzbasierten Lernens möglichst gut zu realisieren. Eine Diskussion dieser Methoden ist für ein interdisziplinäres Verständnis nicht zwingend notwendig, da sie kaum einen Mehrwert gegenüber der noch folgenden Betrachtung der **Stützvektormethoden** bietet.

Einführungsbeispiele

Ein Beispiel für die Umsetzung von menschlichem instanzbasiertem Lernen liegt vor, wenn begonnen wird, ein Puzzle zusammenzusetzen, insbesondere wenn die puzzelnde Person das Bild noch nicht genau betrachtet hat. Üblicherweise wird damit begonnen, einen Puzzlestein zu einem ähnlichen Puzzlestein oder einer zufällig schon bestehenden Häufung von Steinen des entsprechenden Musters zu legen. Dieses Vorgehen wird wiederholt, bis die Steine grob in Kategorien wie Himmel, Rand und so weiter eingeteilt wurden. Die Annahme, dass neue Instanzen aussehen wie bisherige Instanzen, ist in diesem Fall gerechtfertigt, da nur die wenigsten Puzzles Steine enthalten, die keinem anderen Stein ähnlich sehen und jeweils die gesamte Restmenge an Steinen als Referenzgruppe für den isolierten Stein dient. Ein anderes

Beispiel für manuelles instanzensbasiertes Lernen ist das Bleigießen. Hier handelt es sich um ein besonders reines Beispiel, da die konstituierende Vorannahme des Bleigießens genau darin besteht, dass die entstehenden Gebilde bereits bekannten Gebilden ähneln⁴⁶.

Eine grafische Visualisierung für die instanzensbasierte Zerlegung einer Fläche in Teilflächen ist ein **Voronoi-Diagramm**. Die Teilflächen werden in diesem Fall entsprechend der Lage der Trainingsdaten bestimmt. Das jeweils nächstgelegene Trainingsdatum bestimmt die Teilfläche, der ein Punkt zugeordnet wird.

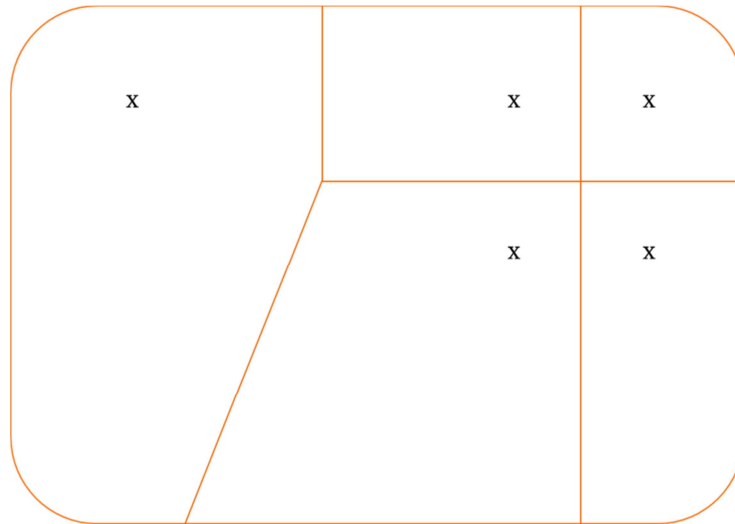


Abbildung 27: Voronoi Diagramm mit fünf Teilflächen

Voronoi-Diagramme veranschaulichen das Resultat der Regel „jede Instanz soll so klassifiziert werden, wie die ihr nächstgelegene Trainingsinstanz“. Die Erstellung des Voronoi-Diagramms selbst ist jedoch gerade kein instanzensbasiertes Lernen, da das Diagramm eine Aussage über den gesamten Hypothesenraum trifft und die Arbeit nach der Erstellung des Diagramms bereits getan ist. Das Diagramm visualisiert lediglich, welches Resultat sich aus instanzensbasiertem Lernen ergibt.

Funktionsbeschreibung

Autoadaptionsprozesse finden im Rahmen von instanzensbasiertem Lernen für eine neu zu klassifizierende Instanz dasjenige bereits klassifizierte Trainingsdatum, das der neuen Instanz am ähnlichsten ist und ordnen beide einer gemeinsamen Klasse zu. Wenn Ähnlichkeit als eine Entfernung interpretiert wird, wird der Raum der Eingabedaten wie im Beispiel des Voronoi-Diagramms in Umgebungen um die Trainingsdaten unterteilt. Instanzensbasiertes Lernen beschreibt lokale Zusammenhänge, die gegebenenfalls nicht für den gesamten Raum der Eingabedaten eine Aussagekraft haben.

Instanzensbasierte Lernstrategien werden als **faule** Lernstrategien bezeichnet, da das Maß zur Bestimmung der Ähnlichkeit vom Nutzer vorgegeben wird und spezielle Trainingsinstanzen erst dann ausgewertet beziehungsweise berücksichtigt werden, wenn eine Eingabe erfolgt, die dies erfordert. Der Autoadaptionsprozess besteht initial nur darin, eine Datenbank der Trainingsdaten zu erstellen. Diese Datenbank stellt in Kombination mit dem Ähnlichkeitsmaß

⁴⁶ Formalere Beispiele und ein technischer Ansatz das instanzensbasierte Lernen insgesamt zu beschreiben, finden sich in einem Text zum „Fallbasierten Problemlösen in Expertensystemen. Begriffliche und inhaltliche Betrachtungen“ (Althoff et Weiß 1991).

formal den Strukturvorschlag dar. Den Strukturvorschlägen anderer Lernvorgänge vergleichbar wäre die Betrachtung von Visualisierungen wie dem Voronoi-Diagramm.

Definition

MLA, die instanzenbasiertes Lernen realisieren, basieren auf zwei Grundsätzen:

- Entscheidungen über die Klassifikation von Eingabedaten werden vorgenommen, wenn konkrete Daten vorliegen.
- Eingabedaten werden auf Ähnlichkeit zu Trainingsdaten überprüft und werden auf Basis des ähnlichsten Teils der Trainingsdaten klassifiziert.

In diesen Grundsätzen ist implizit die Aussage enthalten, dass die weniger ähnlichen Trainingsinstanzen bei der Klassifizierung eines Eingabedatums ignoriert werden und dass keine generalisierenden Hypothesen über die Trainingsdaten aufgestellt werden sollen. Formal wird bei instanzenbasiertem Lernen darüber hinaus gefordert, dass Eingabedaten in einer speziellen mathematischen Codierung vorliegen⁴⁷. Das Voronoi-Diagramm etwa bezog sich auf eine zulässige zweidimensionale Darstellung der Eingabedaten. Die mathematischen Forderungen sind so formuliert, dass eine mathematisch sinnvolle Definition des Abstands der Daten möglich ist und dieser als Maß der Ähnlichkeit verwendet werden kann. Der Abstandsbegriff kann durchaus sehr komplex definiert sein. Im Weiteren wird auf diese Forderung verzichtet, da es für ein interdisziplinäres Verständnis nur einen geringen Mehrwert liefert, zu diskutieren, welche Rohdaten sich auf diese spezielle Weise codieren lassen. Die weiteren Betrachtungen beschäftigen sich dementsprechend formal mit dem **fallbasierten Lernen**. Das fallbasierte Lernen ist ein Spezialfall des instanzenbasierten Lernens, bei dem die Rohdaten relativ frei codiert sein können. Ein Beispiel für fallbasiertes Lernen ist die Erstellung von medizinischen Diagnosen auf Basis der Zuordnung von aufgetretenen Symptomen zu ähnlichen, bereits klassifizierten Symptomen. Zur Bearbeitung von Problemen, die sich gut anschaulich codieren lassen, sich aber gegen eine formale, mathematische Codierung sperren, sind sehr verschiedene Ansätze denkbar. In allen Einzelfällen fließt allerdings sehr viel fallspezifisches Vorwissen in die Konstruktion des Algorithmus ein. Das Ausmaß dieser Anpassungen an die Spezialfälle kann so weit gehen, dass nicht mehr ohne Weiteres davon gesprochen werden kann, dass eine bestimmte Lernstrategie angewandt wird. Im Weiteren wird dennoch nicht zwischen instanzen- und fallbasierten Lernstrategien unterschieden, da die Strategie hinter fallbasiertem Lernen mit derjenigen von instanzenbasiertem Lernen übereinstimmt. Wichtig ist jedoch festzuhalten, dass instanzenbasiertes Lernen durch die Forderung einer bestimmten mathematischen Codierung die Verwendung bestimmter mathematischer Optimierungsverfahren ermöglichen will.

Instanzenbasiertes Lernen kann zwei unterschiedliche Ziele verfolgen, die unterschiedliche Konsequenzen für die Lernstrategie haben. Zum einen kann die Klassifizierung von Eingabedaten gemäß des ähnlichsten Trainingsdatums im Vordergrund stehen. Die alternative Zielvorstellung besteht darin, lokale Hypothesen über die Eigenschaften der Rohdaten zu suchen. Die Rohdaten werden zu diesem Zweck in **Nachbarschaften** zerlegt gedacht, wobei nicht ein Trainingsdatum das Zentrum einer Nachbarschaft darstellt, sondern genau die gegenteilige Idee umgesetzt wird. Für einen Bereich, in dem nicht unmittelbar ein

⁴⁷ Als **reellwertige Punkte** im **n-dimensionalen euklidischen Raum**.

Trainingsdatum vorliegt, werden die nächstgelegenen Trainingsdaten als Nachbarn und der Bereich selbst als eine Nachbarschaft bezeichnet. In der nachfolgenden Abbildung wurde das obige Beispiel um eine Nachbarschaft mit vier Nachbarn erweitert.

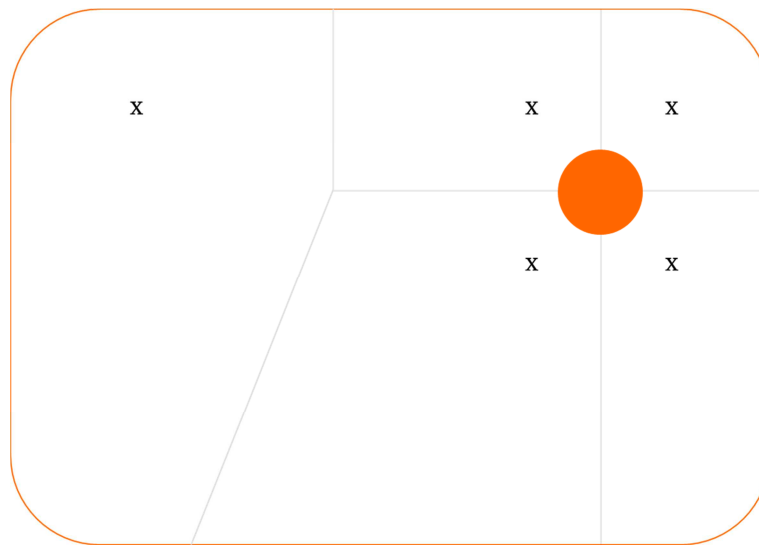


Abbildung 28: Voronoi Diagramm mit einer Nachbarschaft

Die Daten in dieser Nachbarschaft werden als von den vier Nachbarn beeinflusst gedacht. Die Suche nach möglichen Strukturen einer Nachbarschaft bleibt den Grundsätzen des instanzbasierten Lernens treu. Zum einen sollen neue Instanzen ausschließlich unter Nutzung bereits bekannter Instanzen klassifiziert werden, die Ausschließlichkeit bezieht sich insbesondere darauf, dass auf die Formulierung oder Nutzung globaler Hypothesen explizit verzichtet wird. An diese Forderung anknüpfend, soll zum anderen die Komplexität der Strukturvorschläge dynamisch mit Anzahl der Trainingsdaten wachsen – im Gegensatz zur Parametrisierung einer vorgegebenen Menge von Modellen für eine beliebig große Datenmenge. Gleichzeitig wird bei der Betrachtung von Nachbarschaften zwar eine über einen einzelnen Datenpunkt hinausgehende Aussagekraft angestrebt, aber nur insofern als keine Trainingsdaten dem widersprechen. Die speziellen Parameter bei der Erstellung von lokalen Strukturvorschlägen für Nachbarschaften liegen in der Anzahl der zu berücksichtigenden Nachbarn sowie der Wahl und Gewichtung der zu berücksichtigenden Eigenschaften der Trainingsdaten. Die Suche nach lokalen Strukturvorschlägen für Nachbarschaften steht ganz im Sinne des instanzbasierten Lernens im Gegensatz zum **globalen** oder **prädikativen Hypothesenlernen** anderer Lernstrategien. Die Vorgehensweise wird in Abgrenzung von diesem Vorgehen als **lokales** oder **deskriptives Hypothesenlernen** bezeichnet. Auch das lokale Hypothesenlernen ist jedoch wie bereits angedeutet nicht ausschließlich lokal, da zumindest der Abstands- oder Ähnlichkeitsbegriff auf alle Rohdaten anwendbar sein muss.

Varianten instanzbasiertes Lernens

Typische Anwendungssituationen für instanzbasiertes Lernen sind solche, in denen die Nachbarschaften nur lokal definiert werden, sich aber immer mehr verfestigen und entgegen der Grundsätze des instanzbasierten Lernens nicht für jedes zusätzliche Eingabedatum völlig neu berechnet werden sollen. Eine gezielte Verfestigung von Nachbarschaften tritt ebenfalls vergleichsweise häufig auf, etwa wenn eine bestehende und feste Rohdatenmenge weiter analysiert werden soll. In diesem Fall liegt der Fokus nicht darin, Voraussagen für neue

Instanzen zu ermöglichen, sondern die bereits vorliegende Rohdatenmenge besser zu strukturieren. Die beiden in diesem Kontext eingesetzten Varianten instanzenbasierten Lernens sind die der Subgruppenentdeckung und der Clusteranalyse.

Der Fokus der **Subgruppenentdeckung** liegt darauf, besonders interessante Teile der Rohdaten zu identifizieren, wenn die Gesamtmenge an Rohdaten sehr komplex oder chaotisch erscheint. Zu diesem Zweck wird eine Zielfunktion festgelegt, um den Interessantheitsgrad von Gruppen von Rohdaten zu bewerten. Üblicherweise wird die Interessantheit von Rohdaten dabei über eine signifikante Abweichung von einem erwarteten Wert gemessen. Eine solche Abweichung zu messen ist insbesondere dann möglich, wenn die **Verteilung** der Rohdaten zumindest näherungsweise bekannt ist. Als **Subgruppen** werden anschließend entweder diejenigen Gruppierungen von Rohdaten, die einen gewissen Mindestgrad an Interessantheit aufweisen oder die qualitativ besten Gruppen bezeichnet⁴⁸. Zur Veranschaulichung sei angenommen, ein Finanzdienstleister würde planen, zielgruppenorientierte Werbung für die Anschaffung von Kreditkarten zu entwerfen. Von besonderem Interesse sind in diesem Kontext Kundengruppen, die einen strukturellen Zusammenhang aufweisen und unter denen gleichzeitig der Anteil an Kreditkartennutzern relativ zum Gesamtdurchschnitt noch besonders gering ist. Der Finanzdienstleister kann jetzt seine Kundendaten sichten, um entsprechende Subgruppen zu identifizieren und gruppenspezifische Werbemaßnahmen zu initiieren. Bekannt ist in diesem Beispiel lediglich die Zielfunktion. Der Finanzdienstleister muss keinerlei Vorwissen darüber besitzen, welche Gemeinsamkeiten die einzelnen Subgruppen aufweisen. Es muss nicht einmal sichergestellt sein, dass es solche Subgruppen überhaupt gibt. Dieses Beispiel der Subgruppenentdeckung verdeutlicht, dass die Nutzer eines MLA mitunter lediglich vage Wünsche formulieren und das MLA die Aufgabe hat, mittels der Strukturvorschläge Inspirationen für Maßnahmen zur Erfüllung der Wünsche zu liefern. Das MLA kann in dem genannten Beispiel ohne Weiteres bei dem Versuch scheitern, interessante Subgruppen zu entdecken.

Im Rahmen der **Clusteranalyse** werden die Rohdaten so in **Cluster** genannte Teilmengen aufgeteilt, dass Rohdaten innerhalb eines Clusters einander möglichst ähnlich und den Rohdaten außerhalb des Clusters möglichst unähnlich sind. In der Mehrzahl der Fälle ist die Aufteilung in Cluster überschneidungsfrei und bezieht alle Rohdaten mit ein. Zwar wird meist eine Menge von Rohdaten in Cluster aufgeteilt, allerdings kann auch eine **agglomerative Clusteranalyse** durchgeführt werden, bei der Einzeldaten zu kleinen Clustern zusammengefasst werden, die wiederum zu größeren Clustern zusammengefasst werden⁴⁹. Die benötigte Ähnlichkeitsfunktion der Clusteranalyse basiert wie häufig bei instanzenbasiertem Lernen auf einem Abstandsbegriff. Die Ähnlichkeit kann jedoch auch durchaus als eine semantische Ähnlichkeit definiert sein. Dieser Begriff von Ähnlichkeit ist formal vergleichbar mit dem Konzept des Interessantheitsgrades der Subgruppenentdeckung, verzichtet aber gezielt auf eine Hierarchie, da es gerade keine Instanz geben soll, die ein

⁴⁸ Weiterhin werden bei einer Subgruppenentdeckung für instanzenbasiertes Lernen vergleichsweise komplexe Codierungen für Strukturvorschläge zugelassen – etwa indem zeitliche Zusammenhänge zwischen einzelnen Instanzen mitbetrachtet werden.

⁴⁹ Die Aufteilung einer Gesamtheit von Rohdaten hat den Nachteil, dass der Idee des langsamen Anwachsens der Komplexität der Strukturvorschläge entgegengewirkt wird. Im Extremfall führt dieses Vorgehen dazu, dass eine globale Aussage über die Struktur der Gesamtheit der Rohdaten gesucht wird und die Cluster einzelne Aspekte dieser Struktur darstellen sollen. Zwar stellt dieser Extremfall kein instanzenbasiertes Lernen mehr dar, aber die Ausrichtung der Clusteranalyse wird in der Praxis je nach Kontext und Ausmaß der Verschiedenheit der vermuteten Cluster vorgenommen.

hohes Maß an Ähnlichkeit zu allen anderen Instanzen gleichermaßen besitzt. Jenseits der Definition einer Ähnlichkeitsfunktion, die auch bei der Grundform instanzenbasierten Lernens benötigt wird, sind bei der Clusteranalyse keine Vorgaben notwendig. Die Anzahl der Cluster kann prinzipiell im Vorhinein festgelegt werden, eine solche Vorgabe ist jedoch nicht notwendig. Vorwissen oder Vermutungen bezüglich der Strukturen innerhalb der Rohdaten können stattdessen in der Wahl der Mittel zur Umsetzung einer Clusteranalyse genutzt werden. Die im Weiteren noch diskutierte statistische Clusteranalyse unterteilt oder gruppiert Eingabedaten auf Basis der Annahme gewisser Verteilungen zur Schätzung **verborgener Parameter**, das heißt von Parametern, die nicht direkt gemessen werden können. Andere Methoden der Clusteranalyse sind die im Rahmen der Betrachtung von KNN bereits dargestellte adaptive Resonanztheorie und die selbstorganisierenden Merkmalskarten. Diese Beispiele zeigen das Bestehen von Übergangsbereichen zwischen den Lernstrategien an, allerdings verletzen die MLA in diesen Übergangsbereichen meist immer stärker die Grundideen der einen Lernstrategie, wenn sie sich der Vorgehensweise einer anderen Strategie annähern. Die Nutzung von Vorwissen und die gezielte Erstellung desselben führt häufig dazu, dieses als Grundlage des MLA zu verwenden und die Lernstrategien allenfalls als Startpunkt bei der individuellen Gestaltung eines Autoadaptionsprozesses zu betrachten. Wichtig ist festzuhalten, dass auf Vorwissen basierende MLA häufig kaum noch Aspekte aufweisen, die in der Diskussion der Selbstorganisation von Algorithmen eine Rolle spielen.

Darstellungskraft

Die meisten Einsätze von MLA, die auf der Idee des fallbasierten Lernens aufbauen, nutzen in der Umsetzung eine weitere Lernstrategie – die allerdings stark angepasst wird. Der hybride Charakter dieser Teilklasse des instanzenbasierten Lernens ist so ausgeprägt, dass das fallbasierte Lernen als **nachgeordnete Lernstrategie** betrachtet werden kann. Nach Wahl einer Lernstrategie, etwa der KNN, kann entschieden werden, ob ein **instanzenbasierter** oder **konzeptorientierter** Autoadaptionsprozess eingesetzt werden soll. Häufig wird auch innerhalb einer anderen Lernstrategie ein Autoadaptionsprozess auf Basis der Grundsätze des instanzenbasierten Lernens durchgeführt, wenn das Ziel darin besteht, unüberwachtes Lernen zu realisieren.

Instanzenbasierte Lernstrategien erstellen keinen globalen Klassifikator, daher treten auch nicht die damit einhergehenden Einschränkungen bei der Erstellung eines Strukturvorschlages auf. Natürliche Vorgänge und Zusammenhänge sind häufig extrem komplex und lassen sich allenfalls lokal durch einen handhabbaren Strukturvorschlag darstellen. Die Aufteilung in Nachbarschaften erlaubt es instanzenbasierten Lernstrategien, auch chaotisch erscheinende Systeme zu untersuchen und zumindest für Teilsysteme Strukturvorschläge zu erstellen. Entsprechend eignen sich Situationen, in denen kein oder kaum Vorwissen besteht, vergleichsweise gut für instanzenbasierte Lernstrategien, da diese auch dann Muster oder Strukturen finden können, wenn tatsächlich keine globale Struktur vorliegt und der Nutzer im Vorfeld keinen Anhaltspunkt hat, welche Teile der Rohdaten interessante Subgruppen sein könnten. Der korrespondierende Nachteil liegt darin, dass ein Strukturvorschlag, der auf der Suche nach Ähnlichkeiten zwischen Trainingsdaten basiert, keine verborgenen Parameter darstellen kann, da diese sich nicht aus der Registrierung der Eingabedaten ergeben. Dieser Verzicht auf eine Verallgemeinerung der Strukturen der Trainingsinstanzen bedingt, dass Überanpassung ein zentrales Problem des instanzenbasierten Lernens darstellt.

Adaptive Struktur gegen Überanpassung

Überanpassung scheint notwendig gefordert zu sein, wenn alle neuen Instanzen analog zu bereits bestehenden Instanzen klassifiziert werden sollen und gleichzeitig keine neuen Klassen oder etwa Cluster erzeugt werden dürfen. Wenn die Trainingsinstanzen oder Cluster miteinander bezüglich der Einordnung einer neuen Instanz in Konkurrenz stehen, das heißt, wenn **kompetitives** Lernen oder **Wettbewerbslernen** eingesetzt wird, oder wenn für jede neue Instanz nur genau ein Cluster aktualisiert wird, verschärft sich dieses Problem noch. Im Gegenzug kann die Gefahr einer Überanpassung auch reduziert werden, indem diese Form des Lernens nicht eingesetzt wird. Die Alternative besteht darin, neue Instanzen gemäß einer ganzen Nachbarschaft von Trainingsdaten beziehungsweise Clustern zu klassifizieren und die Einflüsse der unterschiedlichen Trainingsdaten gewichtet zu berücksichtigen. So kann eine Vielzahl von Nachbarn berücksichtigt und etwa in Abhängigkeit von ihrer Entfernung gewichtet zur Einordnung der neuen Instanz eingesetzt werden. Einer Überanpassung an eklatante Messfehler oder andere **Ausreißer** in den Rohdaten kann vorgebeugt werden, indem die Umgebungen der Ausreißer bei Überschreitung einer gewissen Entfernung zu den übrigen Trainingsdaten als eigene Nachbarschaften beziehungsweise sehr kleine Einzelcluster interpretiert werden. Die Messung der Performanz auf **Testdaten** und **Validierungsdaten** zur Reduzierung der Überanpassung bietet sich bei instanzenbasierten Lernstrategien besonders an, da im Vorfeld des Autoadaptionprozesses ohne Probleme Trainingsdaten zu Testdaten umgewidmet werden können. Die Trainingsdaten werden im Autoadaptionprozess ohnehin erst genutzt, wenn ein neues Eingabedatum eingeordnet werden soll. Testdaten können bei instanzenbasierten Lernstrategien auch genutzt werden, um optimale Werte für Verfahrensparameter wie die Anzahl der Cluster zu bestimmen oder um diejenigen Attribute der Rohdaten zu identifizieren, die im jeweils verwendeten Abstandsbegriff eine besonders große oder kleine Rolle spielen sollen. Eine gezielte Vernachlässigung einzelner Attribute der Rohdaten ist mitunter sehr zentral für den Autoadaptionprozess. Dies gilt insbesondere, wenn die Rohdaten nicht gut verstanden werden, der Nutzer also im Vorfeld nicht einschätzen kann, welche Attribute überhaupt von Bedeutung sein könnten und dementsprechend eine Vielzahl von Messwerten an das MLA übergibt.

Beispiele für Adaptionen instanzenbasierter Lernstrategien

Im bisher Beschriebenen wurden bereits Veränderungen instanzenbasierter Lernstrategien angedeutet, die in Übergangsbereiche zu anderen Lernstrategien führen. Natürlich kann auch eine Vielzahl von anderen Veränderungen vorgenommen werden, allerdings haben diese meist mit den schon genannten Vorgehensweisen gemein, dass sie den Autoadaptionprozess von den Grundansätzen des instanzenbasierten Lernens entfernen. Nachfolgend wird eine kleine Auswahl an Weiterentwicklungen dargestellt, die einen Eindruck geben soll, auf welche Weise ergänzende Selbstorganisationsprinzipien in den Autoadaptionprozess integriert werden können.

Mitunter sollen einzelne oder ganze Gruppen von Trainingsdaten bewusst nicht in den Lernvorgang einbezogen werden, etwa wenn Messgenauigkeiten vorliegen beziehungsweise vermutet werden. Dies führt zu Veränderungen, die wieder gut in einem Voronoi-Diagramm visualisiert werden können.

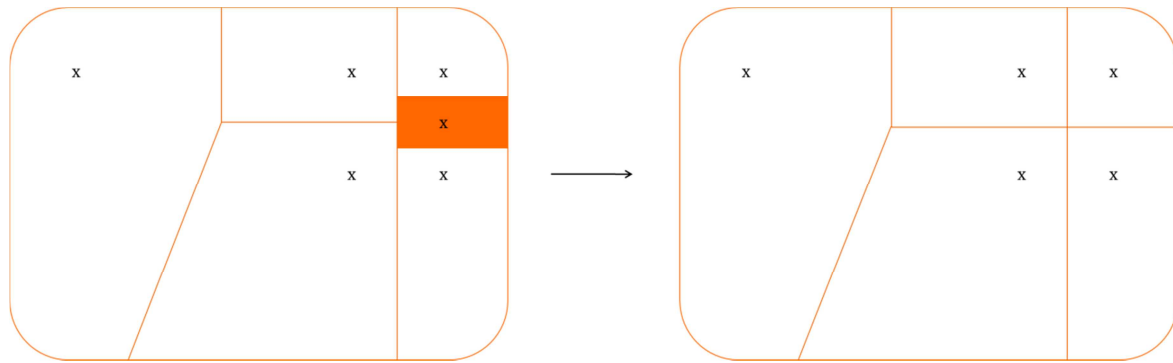


Abbildung 29: Vereinfachung eines Voronoi Diagramms

Alternativ kann auch aus mehreren Trainingsdaten ein künstliches Datum konstruiert werden, allerdings widersprechen beide Maßnahmen gleichermaßen dem grundlegenden Ziel instanzbasierten Lernens, dass keine Trainingsdaten vernachlässigt werden sollen.

Die Subgruppenentdeckung kann verbessert werden, indem bereits gefundene Subgruppen später im Autoadaptionsprozess noch einmal untersucht werden – meist auf zeitliche Zusammenhänge. Ähnliche Subgruppen können anschließend gruppiert werden und der Grad der Interessantheit einer Subgruppe kann noch einmal eingeschätzt werden, diesmal im Vergleich mit ähnlichen Subgruppen und nicht gegenüber der Gesamtmenge an Rohdaten. Wenn eine Subgruppe bezüglich einer anderen keinen großen Informationsgewinn zeigt, kann eine der beiden Gruppen verworfen werden. Der Nutzer wird in diesem Fall lediglich in Form eines Zusatzes zum Strukturvorschlag der verbliebenen Subgruppe über das Verwerfen informiert. Eine solche Vorgehensweise sorgt dafür, dass formal überraschende oder unvorhergesehene, aber dennoch immer wieder vorkommende Strukturen dem Nutzer nicht beliebig oft in ähnlicher Form als neue Strukturvorschläge präsentiert werden. In der weiteren Diskussion wird es von Bedeutung sein, eine grobe Vorstellung zu besitzen, wie MLA eigenständig die Interessantheit von Strukturen bewerten. Die Bewertung identifizierter Subgruppen kann als Beispiel für solch einen Vorgang dienen und ist dementsprechend auch über den Kontext des instanzbasierten Lernens hinaus von Interesse.

Clusteranalysen schließlich müssen keine überschneidungsfreien Cluster generieren und müssen den Raum auch nicht vollständig in Cluster aufteilen. Beispielsweise sind bei **hierarchischen Clusteranalysen** ineinander verschachtelte Cluster zugelassen. Solche vielfach ineinander verschachtelten Cluster können vom Nutzer als Baumstrukturen interpretiert werden und so entsteht ein Hybridfeld zu den Entscheidungsbäumen oder zumindest eine neue Perspektive auf diese Lernstrategie.

2.3.6. Statistisches Lernen

Motivation

Statistische Lernstrategien basieren auf der Arbeit mit und der Manipulation von Wahrscheinlichkeiten. Im Kontext des maschinellen Lernens bedeutet das, dass der Autoadaptionsprozess eine Struktur vorschlagen soll, die mit besonders großer Wahrscheinlichkeit auch über die Trainingsdaten hinaus von Interesse ist. Die Wahrscheinlichkeitsrechnung und die mathematische Stochastik stellen hierzu ein breites Sortiment an sehr stark ausgearbeiteten Begriffsdefinitionen und Werkzeugen zur Verfügung. Die Ziele dieser mathematischen Theoriebildung, liegen in der Prüfung einer Vermutung bezüglich

einer unzugänglichen Gesamtmenge von Rohdaten anhand einer **repräsentativen Stichprobe**. Diese Beschreibung entspricht sehr genau den Anforderungen des maschinellen Lernens. Dementsprechend liegt es nahe, aus wahrscheinlichkeitstheoretischen Werkzeugen statistische Lernstrategien gewinnen zu wollen, die belastbare Aussagen über die Wahrscheinlichkeit der Richtigkeit von Hypothesen im Kontext des maschinellen Lernens machen können.

Analog zu der Betrachtung instanzenbasierter Lernstrategien kann die Diskussion statistischer Lernstrategien dazu beitragen das Verständnis anderer Lernstrategien, die Wahrscheinlichkeiten nicht explizit aber sehr wohl implizit verwenden, zu erweitern. Die später beschriebenen Bayes'schen Lernstrategien ähneln etwa in der Praxis den Ideen des instanzenbasierten Lernens. Entsprechend werden die Begriffe Subgruppenerkennung und Clusteranalyse wieder auftauchen und es wird die Perspektive des statistischen Lernens auf hybride Autoadaptionsprozesse dargestellt werden.

Einführungsbeispiel

Das meistgenutzte Beispiel für ein **Zufallsexperiment** ist der Münzwurf⁵⁰. Statistisches Lernen besteht in dieser Situation lediglich darin, die Münze zu beobachten und nach jedem Wurf neu die Wahrscheinlichkeit zu bestimmen, mit der bisher Kopf geworfen wurde. Der Strukturvorschlag würde dennoch nicht ausschließlich darin bestehen, eine Wahrscheinlichkeit anzugeben, sondern würde zusätzlich eine **Verteilung** prognostizieren und eine Verlässlichkeit der Verteilung angeben. Die Verteilung eines Münzwurfes ist zwar denkbar einfach, aber es wäre aus Perspektive des MLA prinzipiell durchaus möglich, dass zukünftig nicht nur Kopf und Zahl als Eingabedaten registriert werden. Relevant ist die Vorstellung einer auf dem Rand zum Liegen kommenden Münze, weil MLA mögliche implizite Annahmen der Nutzer gerade vermeiden sollen und können und häufig eine Wahrscheinlichkeit dafür abschätzen können, dass konkrete Strukturvorschläge auch bei künftigen Eingabedaten zutreffen werden. Eine Münze wird zwar in der Praxis nicht auf dem Rand landen, aber sie kann durchaus systematisch unterschiedlich oft Kopf und Zahl zeigen. Lediglich eine steigende Zahl von Münzwürfen kann die Verlässlichkeit einer Aussage über die Verteilung der Ergebnisse erhöhen. Allerdings ist auch die Wiederholung eines Zufallsexperimentes nicht unbedingt eine Verbesserung. In dem Fall, dass die Erstellung einer Wahlprognose auf Basis einer Telefonbefragung angestrebt wird, wird die Verlässlichkeit der Ergebnisse sinken, wenn dieselben Personen wieder und wieder angerufen werden.

Funktionsbeschreibung

Statistische Lernstrategien suchen Wege zur Entscheidungsfindung unter expliziter Berücksichtigung und Berechnung oder **Schätzung** der Ungewissheit von Faktoren wie dem Vorwissen oder der Datenerhebung. Statistische Lernstrategien suchen typischerweise Strukturen in Datenbanken und dabei häufig Abhängigkeiten zwischen den Attributen der Rohdaten. Statistische Lernstrategien werden in unterschiedlichen Funktionen eingesetzt. Sie können verborgene Variablen oder Strukturen wie **Verteilungen** suchen oder die Parameter von bereits bekannten oder vermuteten Verteilungen bestimmen. Rohdaten werden dabei immer als Stichproben interpretiert, die genommen wurden, um eine Vermutung zu bestärken oder zu schwächen.

⁵⁰ Ein technischeres Beispiel zur Versicherung von Automobilen findet sich bei Dugas (Dugas et al. 2003).

Statistische Lernstrategien kommen in der Praxis des maschinellen Lernens sehr häufig zum Einsatz und können in Szenarien, in denen andere Lernstrategien besser geeignet erscheinen, als Maßstab dienen, an dem die Performanz anderer Lernstrategien gemessen werden kann.

Definition

Eine **Statistik** ist eine Reihe auf Basis einer **Stichprobe** zu einem **Zufallsexperiment** berechneter beziehungsweise geschätzter Werte. Der wichtigste Begriff in diesem Zusammenhang ist derjenige der Verteilung. Dieser Begriff wurde in Abschnitt 2.2.1 bereits motiviert und verwendet, soll hier aber noch einmal etwas detaillierter dargestellt werden. Eine Verteilung ist ein Graph, der die einem Histogramm zugrunde liegenden Gesetzmäßigkeiten abbildet und für jedes mögliche Ergebnis des Zufallsexperiments die entsprechende Wahrscheinlichkeit angibt. Die Verteilung ignoriert die Schwankungen, die in der Realität entstehen, wenn ein Zufallsexperiment durchgeführt wird. Die wichtigste Verteilung in der Praxis und in der mathematischen Theorie ist die **Normalverteilung**. Diese Verteilung beschreibt jedes Zufallsexperiment, das von vielen unabhängigen Faktoren beeinflusst wird und für das eine große Stichprobe genommen wurde. Nachfolgend ist das Histogramm für eine fiktive Befragung zur Ermittlung der Schuhgröße deutscher Männer abgebildet, die einem normalverteilten Zufallsexperiment entspricht.

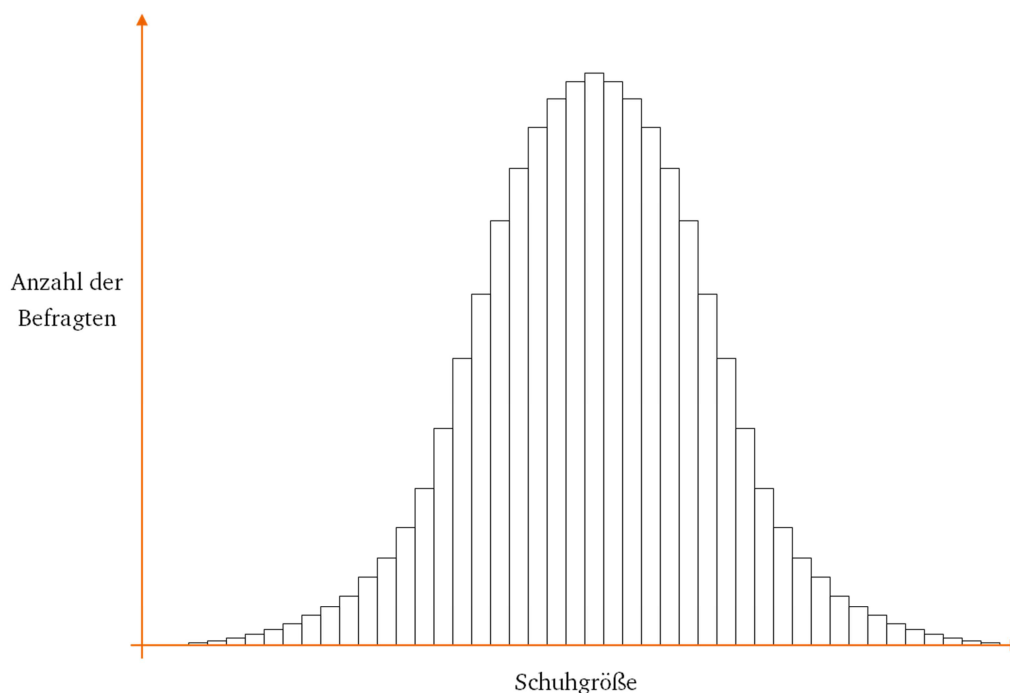


Abbildung 30: Histogramm in Form einer Normalverteilung

Diese fiktive Umfrage ergibt genau die Form einer Normalverteilung, allerdings ist es wichtig zu verstehen, dass eine andere Verteilung angesichts konkreter Trainingsdaten durchaus wahrscheinlicher sein kann als die tatsächlich vorliegende Struktur. Eine Stichprobe ist nicht zwangsläufig repräsentativ für den Kontext, aus dem sie entstammt.

Die Erstellung eines Strukturvorschlages im Rahmen eines Autoadaptionprozesses entspricht einer Entscheidung auf Basis einer Statistik.⁵¹ Solche Entscheidungsfindungen sind mittels

⁵¹ Diese Art der Entscheidung wird formal als **statistische Inferenz** bezeichnet.

parametrischer, semiparametrischer und nichtparametrischer Ansätze möglich. Die Differenzierung dieser Ansätze wurde im Überblick zur Unterscheidung von Algorithmen bereits für das gesamte Feld des maschinellen Lernens erläutert. Der häufigste praktische Einsatz parametrischer oder nichtparametrischer Ansätze findet sich im statistischen Lernen. Statistisches Lernen mittels **parametrischer Ansätze** geht davon aus, dass die dem Zufallsexperiment zugrunde liegenden Verteilungen im Vorfeld bereits bekannt sind. Das bedeutet, in parametrischen Lernszenarien müssen die **Parameter** bestimmt werden, die bekannte Verteilungen auf das konkrete Zufallsexperiment anpassen⁵². Beispiele für solche Parameter sind der **Erwartungswert** und die **Varianz** einer Verteilung⁵³. Ein Beispiel für einen parametrischen Ansatz ist die Methode der maximalen Wahrscheinlichkeit, bei der für eine festgelegte Verteilung und eine konkrete Stichprobe diejenigen Parameter ermittelt werden, die der Entstehung einer Stichprobe der beobachteten Art die größte Wahrscheinlichkeit zuordnen. Die Anwendung dieser Methode verlangt in Konsequenz, dass sämtliche denkbaren Parameter getestet werden und ist dementsprechend aufwendig. Diese Vorgehensweise steht im deutlichen Gegensatz zu dem Ansatz, eine Güte- oder Fitnessfunktion zu definieren oder die Interessantheit einer Struktur zu bewerten, um im Autoadaptionsprozess die jeweils optimalen Parametern zu finden. **Semiparametrische Ansätze** setzen ein Vorwissen zu einem Teil der auftretenden Verteilungen voraus. Ein Beispiel für einen semiparametrischen Ansatz ist eine Clusteranalyse, bei der für manche Cluster von einer parametrisierbaren Verteilung ausgegangen wird und für andere nicht. Die **nichtparametrischen Ansätze** schließlich gehen davon aus, dass die Verteilungen im Vorfeld nicht bekannt sind und als Teil des Strukturvorschlages gesucht beziehungsweise erstellt werden müssen. Ein Autoadaptionsprozess wird auch dann als nichtparametrisch bezeichnet, wenn die Stichprobe nicht mittels einer Verteilung erklärt werden kann. Ein Beispiel für einen nichtparametrischen Ansatz liegt vor, wenn angenommen wird, dass keine systematische parametrisierbare Verteilung vorliegt, sondern nur die Eintrittswahrscheinlichkeiten unabhängiger Einzelergebnisse berechnet beziehungsweise abgeschätzt werden. Parametrische und nichtparametrische Ansätze werden zwar auf unterschiedliche Weise im Rahmen der Autoadaptionsprozesse realisiert, die Unterschiede liegen allerdings in erster Linie in der Wahl der eingesetzten mathematischen Methoden. Die prinzipiellen Vorgehensweisen sind jeweils vergleichbar. Die exemplarische Betrachtung einer Variante statistischen Lernens ist daher ausreichend, um ein prinzipielles Verständnis des statistischen Lernens als einem Teilbereich des maschinellen Lernens zu erlangen. Die folgenden Darstellungen werden sich entsprechend auf das **Bayes'sche Lernen** konzentrieren. Bayes'sches Lernen ist einerseits in der Praxis von großer Relevanz und basiert andererseits auf einer sehr kompakten mathematischen Grundlage. Die Betrachtung des Bayes'schen Lernens ermöglicht die Entwicklung einer nützlichen Intuition, wie mathematisches Hintergrundwissen im maschinellen Lernen zum Einsatz kommen kann. Im Weiteren wird zwischen der Grundform und einer modellbasierten Variante Bayes'schen Lernens unterschieden – zwischen dem **Assoziationslernen** und den **Bayes'schen Netzen**.

⁵² In der Statistik wird an dieser Stelle der Begriff **Schätzer** benötigt, um diejenigen Parameter zu benennen, die bestimmt werden müssen, bevor wiederum Aussagen zu den, häufig verborgenen, Parametern der Verteilungen möglich sind.

⁵³ Der Erwartungswert beschreibt den Wert, der als mittlere Ausgabe erwartet werden kann, wenn mittels der Verteilung ein zufälliger Wert ausgegeben wird. Die Varianz beschreibt, wie stark die Verteilung um den Erwartungswert herum variiert, das heißt, wie sehr die Werte der Verteilung voneinander abweichen können.

Bayes'sches Lernen

Bayes'sches Lernen basiert wie die meisten Varianten statistischen Lernens auf der Arbeit mit einem speziellen Teil der mathematischen Begriffswelt und mit speziellen Methoden zum Umgang mit Wahrscheinlichkeiten. Der in diesem Fall zum Einsatz kommende und für die mathematische Theorie sehr grundlegende Begriff ist derjenige der **bedingten Wahrscheinlichkeit**. Angenommen, ein Biosupermarkt führt eine Befragung von Kunden durch und erkundigt sich, ob die befragte Person sich vegetarisch ernährt. Der Anteil der Kunden, die diese Frage bejahen, könnte nun sehr stark von Aspekten wie dem Bio-Sortiment des Supermarktes beeinflusst werden. Mit einer solchen Umfrage wird die Wahrscheinlichkeit gemessen, dass eine Person sich vegetarisch ernährt, unter der Bedingung, dass sie in diesem speziellen Biosupermarkt einkauft. Diese Statistik wird als eine bedingte Wahrscheinlichkeit bezeichnet. Die bedingte Wahrscheinlichkeit eine Person zu interviewen, die sich vegetarisch ernährt, unterscheidet sich je nach Kontext der Interviewreihe möglicherweise erheblich.

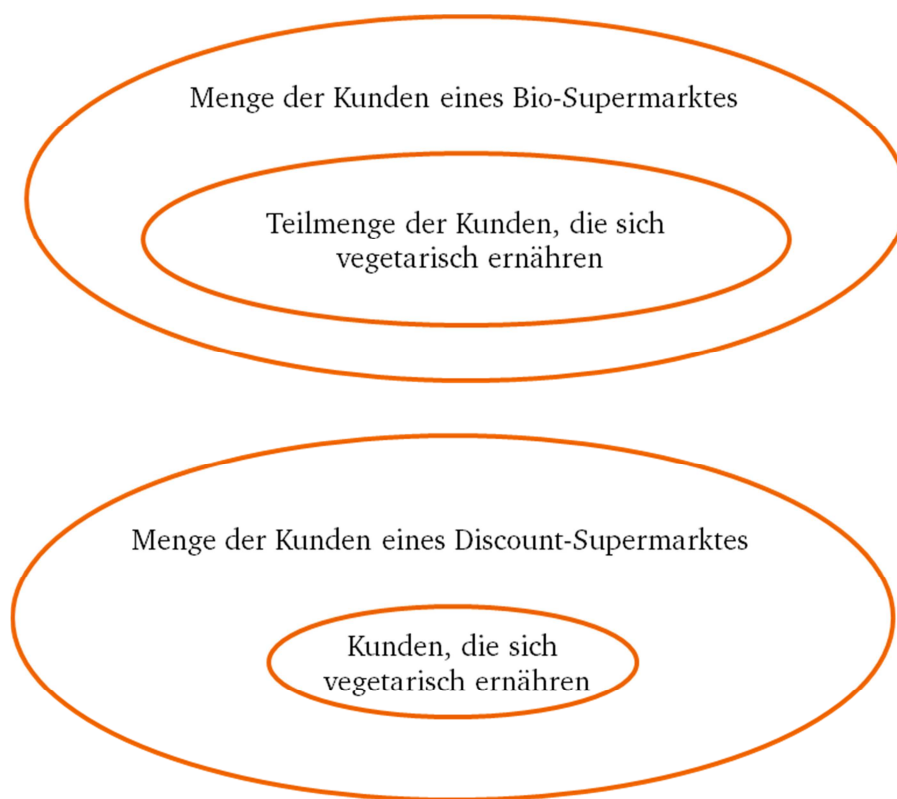


Abbildung 31: Bedingte Wahrscheinlichkeit des Antreffens von Vegetariern

Die Bestimmung einer bedingten Wahrscheinlichkeit unterscheidet sich nicht von derjenigen einer **unbedingten Wahrscheinlichkeit**. Im Falle der Messung einer unbedingten Wahrscheinlichkeit wird die Häufigkeit eines bestimmten Ereignisses bestimmt und in ein Verhältnis zur Gesamtzahl von Ereignissen gesetzt. Eine bedingte Wahrscheinlichkeit wird analog ermittelt. Der einzige Unterschied besteht darin, dass die Gesamtzahl von Ereignissen aufgrund einer expliziten Vorannahme reduziert wird und nur eine Teilmenge der Gesamtzahl von Ereignissen betrachtet wird. Die Bedeutung solcher bedingten Wahrscheinlichkeiten ergibt sich wie folgt: sollte die Wahrscheinlichkeit Vegetarier anzutreffen in dem betrachteten Biosupermarkt deutlich größer sein, als es dem Anteil der Vegetarier an der Gesamtbevölkerung entspricht, so kann vermutet werden, dass der Besuch des Biosupermarktes nicht unabhängig von den Essgewohnheiten einer Person ist. Anders gesagt,

bedingt die Wahl des Supermarktes die Essgewohnheiten und die Essgewohnheiten bedingen die Wahl des Supermarktes. Dieser Zusammenhang wird als **Assoziationsregel** bezeichnet und es kann nicht auf das Vorliegen einer Kausalität und insbesondere nicht auf die Richtung einer Kausalität geschlossen werden. Der typische Fehler an dieser Stelle besteht darin, zu vermuten, dass das Betreten eines Bio-Supermarktes Menschen zu einer vegetarischen Ernährung veranlasst. Ein Beispiel für die Aufklärung solch einer Fehlinterpretation besteht darin, dass eine Langfriststudie über mehrere Jahrzehnte durchgeführt wurde, um die Auffassung zu widerlegen, dass Menschen aufgrund einer vegetarischen Ernährung eine höhere Lebenserwartung besitzen (Chang-Claude et al. 2005).

Bedingte Wahrscheinlichkeiten sind im statistischen Lernen von großer Bedeutung, weil bei der Suche nach der Hypothese, die das Ergebnis einer speziellen Stichprobe am wahrscheinlichsten erklärt, in erster Linie solche bedingten Wahrscheinlichkeiten bestimmt werden müssen. Die mathematische Theorie reduziert mit dem **Satz von Bayes** für die Suche nach der wahrscheinlichsten Hypothese die Anforderungen auf zwei Statistiken. Erstens muss die unbedingte Wahrscheinlichkeit der Hypothese selbst bestimmt werden und zweitens muss die bedingte Wahrscheinlichkeit bestimmt werden, dass die Trainingsdaten auftreten – unter der Bedingung, dass die Hypothese richtig ist⁵⁴.

Praktische Anwendungen Bayes'schen Lernens finden sich bei vielen Varianten von Produktempfehlungen. Eines der bekanntesten Beispiele ist der Onlineversand von Amazon. Dem Kunden wird bei Amazon zu jedem Kauf vorgeschlagen, welche Artikel von anderen Kunden zusammen mit den vom ihm selbst gekauften Artikeln erstanden wurden. Das heißt, der Lernvorgang besteht in diesem Fall nur darin, zu jedem Artikel eine Liste der ebenfalls gekauften Artikel zu aktualisieren und diese Liste bei Bedarf anzuzeigen. Die dahinterliegende Idee besteht in der Messung einer bedingten Wahrscheinlichkeit. Wenn die Wahrscheinlichkeit des Kaufes eines Artikels unter gewissen Bedingungen höher ist, möchte der Versandhandel die Artikel genau dann bewerben, wenn diese Bedingungen gerade eingetreten sind. Das Vorliegen einer Kausalität spielt hierbei für den Händler keine Rolle und wird auch nicht vorausgesetzt, das Ziel ist der Verkauf eines weiteren Artikels. Generell sind **Warenkorbanalysen** zur Erstellung von Kundenprofilen und zielgruppengerechter Werbemaßnahmen typische Anwendungen für Bayes'sches Lernen. Das Beispiel des Onlineversands hat für den Händler den Vorteil, dass dort eine erfolgreiche Werbemaßnahme unmittelbar den Umsatz steigert und eine fehlgeschlagene Werbemaßnahme annähernd ohne Konsequenzen bleibt. In Kontexten hingegen, in denen medizinische Verträglichkeitsuntersuchungen durchgeführt werden oder die Verlässlichkeit von Diagnosen geprüft werden soll, sind fehlerhafte Strukturvorschläge sehr viel folgenreicher. Wichtig ist hier festzustellen, dass die Verfügbarkeit einer ausgearbeiteten mathematischen Theorie in der Praxis eine große Stärke statistischen Lernens darstellt, da beispielsweise Zuverlässigkeitsaussagen bezüglich der Strukturvorschläge möglich sind. Die Darstellung der beiden Hauptvarianten Bayes'schen Lernens bietet jedoch bereits genügend Gelegenheit zur Entwicklung eines interdisziplinären Verständnisses statistischen Lernens. Die Betrachtung des statistischen Lernens wird daher wie angekündigt im Rahmen der Betrachtung des Assoziationslernens und der Bayes'schen

⁵⁴ Nicht benötigt wird etwa die bedingte Wahrscheinlichkeit, dass die Hypothese richtig ist, falls die Trainingsdaten aufgetreten sind. Der aus der mathematischen Theorie folgende abduktive Bias liegt darin, dass von der Korrektheit dieser beiden Statistiken ausgegangen werden muss und dass beide zumindest teilweise aus Vorwissen berechnet werden müssen.

Netze erfolgen und es wird keine Diskussion der kontextabhängigen mathematischen Weiterentwicklungen angestrebt.

Assoziationslernen

Das Ziel des Autoadaptionprozesses beim **Assoziationslernen** besteht in der Suche nach Zusammenhängen in Form von bedingten Wahrscheinlichkeiten, den **Assoziationsregeln**⁵⁵. Die gesuchten bedingten Wahrscheinlichkeiten zu berechnen ist prinzipiell sehr einfach, da lediglich Häufigkeiten verglichen werden müssen. Die Herausforderung ergibt sich daraus, dass Assoziationsregeln Aussagen zu beliebig vielen der Attribute der Daten beinhalten können und es daher extrem viele mögliche Assoziationsregeln gibt. Angenommen, im Beispiel des Biosupermarktes ist wiederum die Wahrscheinlichkeit gesucht, dass ein Kunde sich vegetarisch ernährt. In diesem Fall könnte die Bedingung der Wahrscheinlichkeit sein, dass nur Kunden betrachtet werden, die Vormittags befragt wurden oder dass darüber hinaus nur diejenigen Personen relevant sind, die Linkshänder sind und zum Zeitpunkt des Interviews ein grünes Kleidungsstück getragen haben.

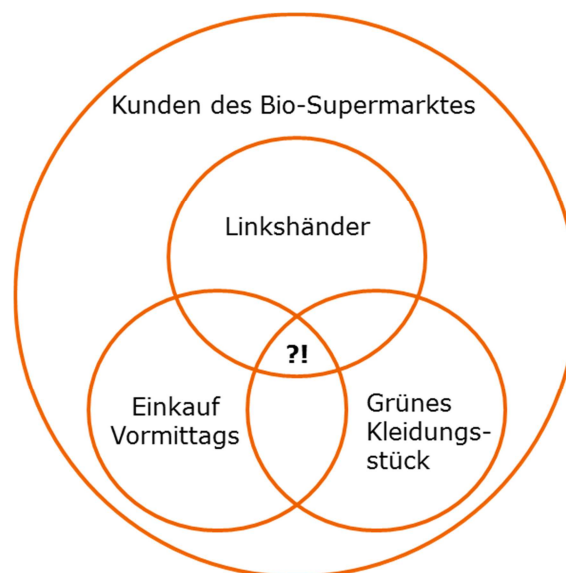


Abbildung 32: Bedingte Wahrscheinlichkeiten mit mehreren Attributen

Hier wurde zuerst ein einzelnes und dann drei der Attribute der Trainingsdaten als Bedingung für die Wahrscheinlichkeit angesehen.

Assoziationsregeln weichen etwas von der Grundidee des statistischen Lernens ab, da alle korrekt berechneten Assoziationsregeln die Trainingsdaten gleich gut abbilden. Die Idee hinter der Erstellung einer Assoziationsregel ist weniger, die Eigenschaften zukünftiger Rohdaten prognostizieren zu können als vielmehr, interessante Eigenschaften der vorliegenden Trainingsdaten zu beschreiben. Wenn die Prognosefähigkeit im Fokus steht, werden entsprechend Methoden des instanzbasierten Lernens mitbetrachtet und es wird versucht, interessante Assoziationsregeln zu identifizieren. Die dafür notwendige Suche nach interessanten Assoziationsregeln ist aufgrund der extrem großen Anzahl von denkbaren Assoziationsregeln sehr kompliziert⁵⁶. Die unüberschaubare Anzahl denkbarer Bedingungen

⁵⁵ Entsprechend werden weder eine Verteilung noch Parameter einer solchen gesucht.

⁵⁶ Die Möglichkeiten eine beliebige Teilmenge aus einer Gesamtmenge auszuwählen ergeben zusammen die Potenzmenge der Gesamtmenge. Diese wächst exponentiell mit der Anzahl der Elemente der Gesamtmenge.

für Wahrscheinlichkeiten wird in der Praxis häufig mit Hilfe von einfachen Vorgaben drastisch reduziert. Beispielsweise kann verlangt werden, dass mindestens zwei Trainingsdaten auf einmal betrachtet werden und dass jede formulierte Aussage mindestens für diese beiden Trainingsdaten zutreffend ist. Die Interessanztheit einer bedingten Wahrscheinlichkeit ergibt sich darüber hinaus nicht rein aus ihrer Größe. So könnte die Zahl der Kunden in der obigen Schnittmenge sehr klein sein und keine Kunden enthalten, die sich vegetarisch ernähren. Die bedingte Wahrscheinlichkeit wäre damit gleich null. Diese Assoziationsregel ist dennoch nicht sehr interessant, wenn die Aussage nur sehr wenig Kunden betrifft. Assoziationsregeln werden daher noch in einer zweiten Dimension bewertet, der Anzahl von Trainingsdaten, die von einer bestimmten Regel noch betroffen werden. Die nachfolgende Visualisierung zeigt ein Beispiel für eine Menge von Trainingsdaten, die als geometrische Objekte dargestellt sind.

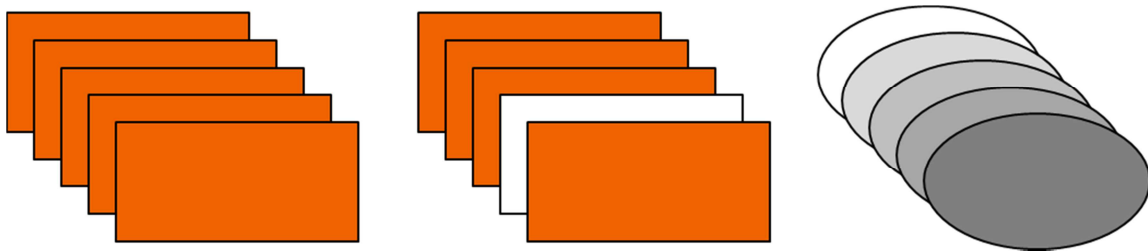


Abbildung 33: Trainingsdaten in Form von geometrischen Objekten

In diesem Beispiel könnte eine Assoziationsregel lauten „unter der Bedingung, dass ein weißes Objekt gewählt wird, ist Wahrscheinlichkeit ein rechteckiges Objekt zu erhalten 50%“. Diese Assoziationsregel betrifft allerdings nur zwei Trainingsdaten. Die Assoziationsregel „unter der Bedingung, dass ein Rechteck gewählt wird, ist die Wahrscheinlichkeit ein orangefarbenes Objekt zu erhalten 90%“ hingegen betrifft zehn Objekte und erreicht dennoch einen hohen Prozentwert. In den meisten Kontexten wäre eine Assoziationsregel der zweiten Art damit interessanter. Wichtig ist hier, dass auch 50% noch eine vergleichsweise große bedingte Wahrscheinlichkeit ist, da insgesamt sechs Farben in den Trainingsdaten vertreten sind und die weißen Objekte mit einem Anteil von zwei von fünfzehn nur 13% der Gesamtzahl von Objekten ausmachen.

Zusammengefasst kann die Suche nach Assoziationsregeln mit Hilfe der folgenden Schritte A bis E beschrieben werden.

- A** Das MLA sucht nach Aussagen, die sehr spezielle Zusammenhänge zwischen einer kleinen Anzahl von Trainingsdaten beschreiben und berechnet die zugehörigen bedingten Wahrscheinlichkeiten.

Aussagen, die Anforderungen an besonders viele Attribute der Trainingsdaten stellen, betreffen meist nur sehr kleine Teilmengen der Trainingsdaten. Entsprechend ergeben sich aus den Trainingsdaten meist vergleichsweise niedrige oder hohe bedingte Wahrscheinlichkeiten. Von Interesse sind im Autoadaptionsprozess die vergleichsweise großen bedingten Wahrscheinlichkeiten, die jedoch sehr wahrscheinlich nur Teile der Trainingsdaten betreffen. Das heißt, es sollte der Grad der Spezialisierung der Aussage gesenkt werden, auch wenn dadurch die bedingte Wahrscheinlichkeit sinkt.

- B** Nach der Identifikation einer bedingten Wahrscheinlichkeit, die größer als ein im Vorfeld festgelegter Grenzwert ist, wird geprüft, ob die zugehörige Aussage eine

Anzahl von Trainingsdaten betrifft, die größer als ein zweiter im Vorfeld festgelegter Grenzwert ist.

- C** Wenn zu wenige Trainingsdaten betroffen sind, werden in der entsprechenden Aussage enthaltene Anforderungen verworfen, bis die Aussage entweder ausreichend viele Trainingsdaten betrifft oder die bedingte Wahrscheinlichkeit nur noch knapp über dem Grenzwert liegt und die Aussage verworfen wird.

Im obigen Beispiel könnte eine Aussage über vormittags einkaufende, linkshändige Personen, die grüne Hosen tragen, zu einer Aussage über vormittags einkaufende, linkshändige Personen werden. Es ist anzunehmen, dass die Menge an Supermarktkunden, die die genannten Merkmale aufweisen, durch die Entfernung der Forderung eines grünen Kleidungsstücks deutlich größer geworden ist. Es kann weiter angenommen werden, dass der Anteil der Menschen, die sich vegetarisch ernähren, sich durch die Vergrößerung der Gruppe stark verändert hat.

- D** Wenn die resultierende Aussage eine Anzahl von Trainingsdaten betrifft, die größer als ein zweiter im Vorfeld festgelegter Grenzwert ist, wird die Aussage als Assoziationsregel bezeichnet, festgehalten und an den Nutzer übermittelt. Wenn die resultierende Aussage sich auf zu wenige Trainingsdaten bezieht, wird sie verworfen.

- E** Der Prozess beginnt wieder mit Schritt A.

Das Verwerfen von Anforderungen ähnelt der Stutzung von Entscheidungsbäumen, wenngleich im statistischen Lernen deutlich andere Schwerpunkte gesetzt werden. Die Idee besteht nicht darin, die Rohdaten in Klassen aufzuteilen und es soll keine grafische Repräsentation erstellt werden. Stattdessen sollen Aussagen über bedingte Wahrscheinlichkeiten getroffen werden und dazu ist es notwendig den Interessantheitsgrad von Aussagen zu bewerten. Die resultierende Suche nach interessanten Assoziationsregeln ähnelt wiederum dem instanzbasierten Lernen und speziell der Subgruppenentdeckung, allerdings ist das Ausmaß dieser Ähnlichkeit veränderlich. Das Assoziationslernen wird in erster Linie über die Vorgabe der beiden genannten Grenzwerte beeinflusst – der Vorgabe einer minimalen Menge von Trainingsdaten, die von der Assoziationsregel betroffen sein müssen und einer minimalen Höhe für die bedingte Wahrscheinlichkeit. Mit Hilfe dieser Grenzwerte kann insbesondere gesteuert werden, ob die Assoziationsregeln auf die Gesamtmenge der Trainingsinstanzen anwendbar sind oder ob ihre Aussagen nur für spezielle Teilbereiche gültig sind. Diese Wahl entscheidet entsprechend, wie sehr sich im Assoziationslernen die Idee der Subgruppenentdeckung widerspiegelt.

Unabhängig davon, wie der Ablauf der Schritte von A bis E durch externe Vorgaben beeinflusst wird, ist es problematisch diese Vorgehensweise als Autoadaptionsprozess zu bezeichnen. Zwar werden Strukturvorschläge ausgegeben und der Prozess orientiert sich an Trainingsdaten, allerdings wird dabei nur in geringem Ausmaß autoadaptiv vorgegangen. Die konkrete Durchführung von Schritt C erfordert ein Vorgehen, das anderen Lernstrategien ähnelt, etwa bezüglich der Festlegung, in welcher Reihenfolge die Anforderungen fallen gelassen werden oder ob Anforderungen graduell oder vollständig fallen gelassen werden. Das Vorgehen erinnert dennoch insgesamt mehr an ein mathematisches Optimierungsverfahren als an ein Konzept, das eine Form von Selbstorganisation abbildet.

Bayes'sche Netze

Bayes'sche Netze nutzen bedingte Wahrscheinlichkeiten, um lokale Zusammenhänge innerhalb der Trainingsdaten nachzubilden und setzen diese lokalen Zusammenhänge im Rahmen des Autoadaptionsprozesses zu einer globalen Struktur zusammen. Bayes'sche Netze basieren auf der Idee, grundlegendes Vorwissen über den Kontext unmittelbar in ihrer Struktur abzubilden und können entsprechend auch von einem unerfahrenen Nutzer leicht erstellt werden. Sogar unklare oder umstrittene Zusammenhänge können vom Nutzer sofort in die Struktur des Netzes integriert werden, so dass diese im Rahmen des Autoadaptionsprozesses geprüft werden können.

Die grafische Darstellung eines Bayes'schen Netzes entspricht einer Ansammlung von Knoten, die über gerichtete Verbindungen, die keine Zirkel bilden dürfen, miteinander verbunden sind. Die Darstellung entspricht der eines **azyklischen Graphen** und ähnelt formal der Darstellung von azyklischen künstlichen neuronalen Netzen. Die Methoden zur Bestimmung der Struktur und zum Umgang mit Bayes'schen Netzen spiegeln dies auch wieder, allerdings dient die grafische Darstellung bei Bayes'schen Netzen der Erweiterung des Verständnisses der Nutzer und spiegelt deren Hintergrundwissen beziehungsweise deren Vermutungen wieder. Die Knoten werden zu diesem Zweck jeweils mit einem **Einflussfaktor**⁵⁷ identifiziert und die möglichen Zustände dieses Einflussfaktors werden in Form einer **Wahrscheinlichkeitstabelle** festgehalten. Am einfachsten kann die Struktur Bayes'scher Netze an einem Beispiel dargestellt werden. Angenommen, es soll die Wahrscheinlichkeit abgeschätzt werden, bei der Überquerung einer Straße in der Nähe einer Kraftfahrzeug-Ampel einen Unfall zu erleiden. Diese Situation lässt sich als bedingte Wahrscheinlichkeit ausdrücken. Die bedingten Wahrscheinlichkeiten lassen sich dabei als Tabelle darstellen, in der neben den drei Ampelfarben⁵⁸ auch zwei unterschiedlich schwere Unfälle und ein unfallfreier Normalfall unterschieden werden. In der Visualisierung wird weiter angenommen, dass die beobachtete Ampel gleich lang ein rotes und grünes Signal zeigt, während nur in 10% der Zeit ein gelbes Signal zu sehen ist.

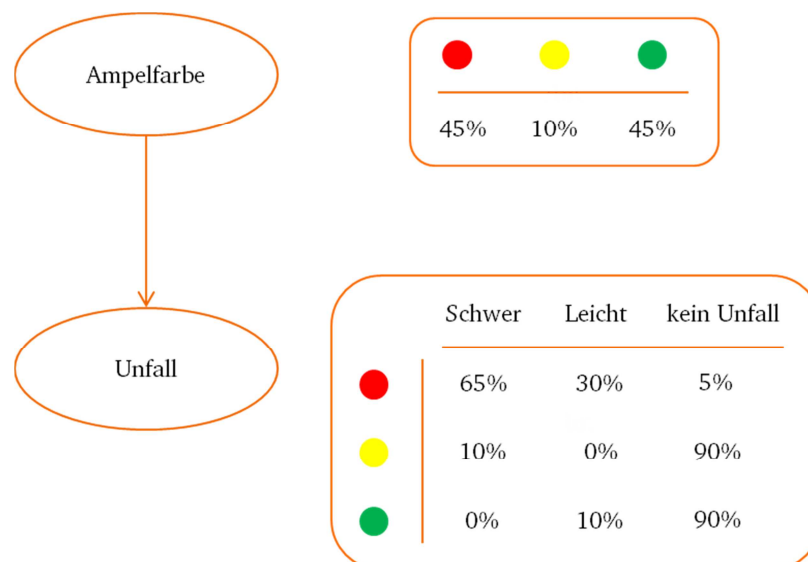


Abbildung 34: Bayes'sches Netz mit Wahrscheinlichkeitstabellen

⁵⁷ Dieser Faktor wird formal als **Zufallsvariable** bezeichnet.

⁵⁸ Es wurde angenommen, dass „Gelb-Rot“ aus Sicht der Fußgänger einem roten Signal entspricht.

Das Ergebnis dieses Netzes wäre die Erkenntnis, dass die Unfallgefahr bei gelbem und grünem Licht gleich groß ist, dass jedoch die Schwere des Unfalls unterschiedlich ist. Basierend auf diesem Netz kann nun geschätzt werden wie wahrscheinlich es ist, dass jemand angefahren wird.

Die gerichtete Verbindung zwischen dem Knoten der Ampelfarbe und dem Knoten des Unfalls gibt an, dass die Ampelfarbe die Schwere des Unfalls beeinflusst, allerdings nicht notwendigerweise kausal erzeugt, da es eine beliebige Anzahl von Einflussfaktoren für jeden Knoten geben kann. Bayes'sche Netze mit nur einem Einflussfaktor, wie es hier die Ampelfarbe war, stellen lediglich bedingte Wahrscheinlichkeiten dar, aber schon die Einbeziehung von einem zweiten Einflussfaktor lässt ein sehr komplexes Ergebnis entstehen. Die Unterscheidung von Tag und Nacht führt für das obige Beispiel zu folgender Struktur.

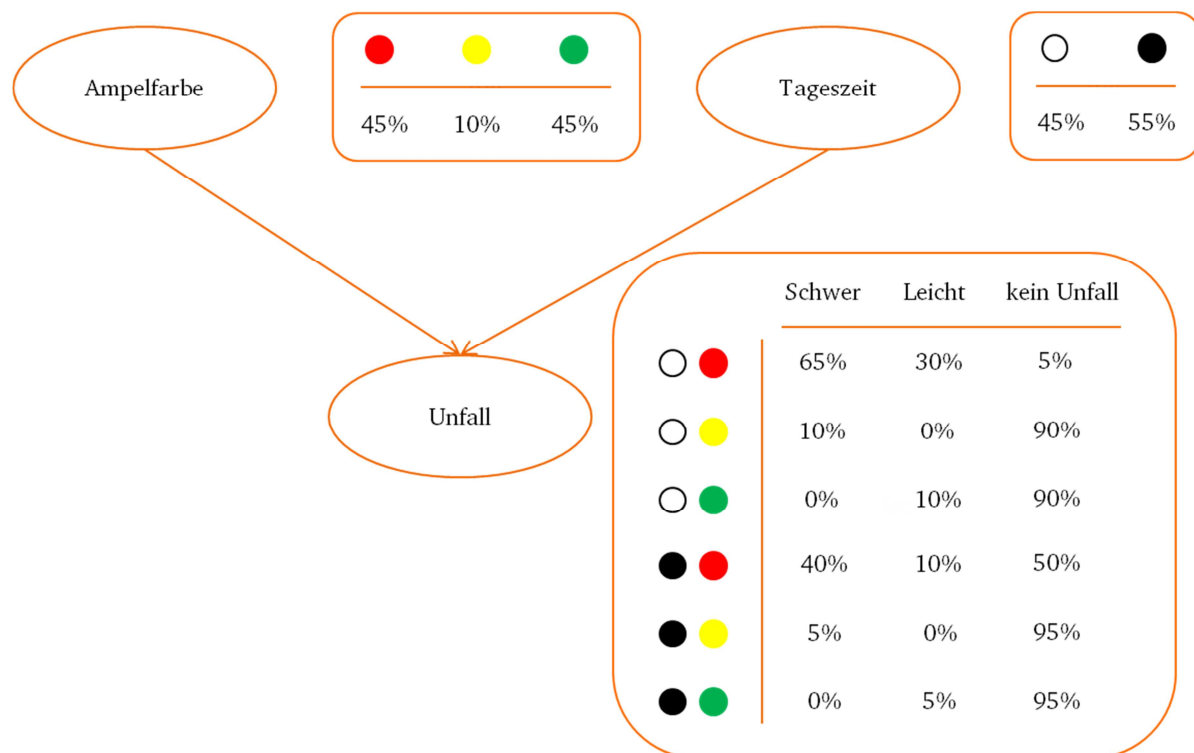


Abbildung 35: Bayes'sches Netz mit zwei Einflussgrößen

In diesem fiktiven Fall ist erkennbar, dass die Messreihe des ersten Beispiels wahrscheinlich tagsüber erstellt wurde und dass die Überquerung einer Straße in der Dunkelheit deutlich sicherer ist als bei Tageslicht. Von den beiden genannten Einflussfaktoren abgesehen, könnte es noch eine Vielzahl von nicht ohne Weiteres messbaren Einflussfaktoren geben, wie das durchschnittliche Verkehrsaufkommen am jeweiligen Kalendertag oder die mittlere Bremskraft eines Kraftfahrzeugs. Solche verborgenen Parameter, deren Existenz aus Vorwissen gefolgert wurde, können genau wie den messbaren Einflussfaktoren direkt zusätzlichen Knoten zugeordnet werden.

Ein praktisches Beispiel für den Einsatz von Vorwissen speziell in der medizinischen Diagnose ist die Annahme der Existenz einer Krankheit, die bestimmte beobachtbare Symptome erzeugt. Diese Krankheit selbst kann zwar nicht gemessen, aber dennoch als Einflussfaktor in ein Bayes'sches Netz aufgenommen werden. In diesem Aspekt grenzen sich Bayes'sche Netze wesentlich gegen das Assoziationslernen ab, bei dem die Darstellung von verborgenen Parametern nicht möglich war. Weiter werden verborgene Parameter im Gegensatz zu KNN

transparent dargestellt und die Erstellung der Ihnen zugeordneten Knoten beruht direkt auf dem Vorwissen des jeweiligen Nutzers.

Das typischste Verfahren zur Realisierung Bayes'scher Netze ist der **Expectation-Maximization-Algorithmus**⁵⁹, ein iterativer Autoadaptionsprozess zur Bestimmung verborgener Parameter, der auf der Vorgabe einer Netzstruktur aus Vorwissen aufbaut⁶⁰. Die Bezeichnung des Algorithmus leitet sich direkt aus seinem Vorgehen ab:

- A** Im Erwartungsschritt werden mittels der aktuell prognostizierten Wahrscheinlichkeitstabellen die zu erwartenden Werte der nicht messbaren Einflussfaktoren berechnet. Diese Parameterwerte werden für den nächsten Schritt behandelt, als wären sie gemessen worden und somit Teil der Trainingsdaten.
- B** Im Maximierungsschritt werden die Wahrscheinlichkeitstabellen den Einflussfaktoren so angepasst, dass die Wahrscheinlichkeit des Auftretens der Trainingsdaten möglichst groß ist.
- C** Anschließend wird wieder mit Schritt **A** fortgefahren, bis ein stabiler Zustand erreicht wird.

Kurz gesagt, wird immer abwechselnd angenommen, dass die angenommenen Wahrscheinlichkeitstabellen respektive die geschätzten Parameter korrekt vorliegen.

Die Hauptstärke Bayes'scher Netze liegt wie bereits beschrieben darin, komplexe Kontexte mittels gut verständlicher Aussagen zu Teilstrukturen zu modellieren. Insofern realisieren Bayes'sche Netze eine **Clusteranalyse**, insbesondere kann der EM-Algorithmus in vielen Kontexten als ein Verfahren zur Clusteranalyse betrachtet werden. Das bedeutet jedoch nicht, dass Bayes'sche Netze einen Teilbereich des instanzbasierten Lernens darstellen, da die Grundidee deutlich in der Nutzung der bedingten Wahrscheinlichkeiten zur verständlichen und intuitiven Darstellung von Zusammenhängen zwischen verschiedenen Einflussfaktoren liegt. Aus der interdisziplinären Perspektive ist festzuhalten, dass hier zwei unterschiedliche Lernstrategien zur Erstellung von Autoadaptionsprozessen durch verschiedene, aber doch vergleichbare Varianten eines Algorithmus realisiert werden, ohne dass die jeweilige Grundidee aufgegeben wird.

Eine Schwäche Bayes'scher Netze, die über das statistische Lernen hinaus von Interesse ist, stellt der Effekt des **Hinwegerklärens** dar. Dieser Effekt äußert sich beispielsweise darin, dass wenn – unabhängig von dem Beispiel der Verkehrssicherheit – bereits *bekannt* ist, dass eine Straße nass ist, die möglichen Ursachen für die Nässe voneinander abhängig sind beziehungsweise werden. Die Wahrscheinlichkeit, dass es geregnet hat, sinkt beispielsweise, wenn zusätzlich bekannt ist, dass gerade jemand sein Auto gewaschen hat. In diesem Fall ist der Fakt, dass die Straße nass ist, bereits erklärt und kann zur Beantwortung der Frage, ob es geregnet hat, nicht mehr unmittelbar genutzt werden. Eine Netzstruktur, die diesen Einfluss

⁵⁹ Kurz **EM-Algorithmus**. Eine technische Darstellung, die mit dem EM-Algorithmus arbeitet und bei Vorliegen sehr guter mathematischer Vorkenntnisse einen interessanten Eindruck vermittelt, findet sich bei Friedman (Friedman 1998).

⁶⁰ Die Netzstruktur selbst kann im Rahmen des Erwartungs-Maximierungs-Algorithmus ebenfalls adaptiert werden, indem mehrere Strukturen verglichen werden. Die Anzahl der möglichen Netzstrukturen wächst jedoch extrem schnell mit der Anzahl der Knoten, so dass es nicht möglich ist, alle zu überprüfen. Die Gegenmaßnahme hierzu besteht darin, dass die Einflussfaktoren im Gegensatz zu KNN nur dann über gerichtete Verbindungen mit anderen Einflussfaktoren beziehungsweise Knoten verbunden werden, wenn auch wirklich ein Zusammenhang vermutet wird.

der Autowäsche auf das Wetter abbildet, ist weder falsch noch unplausibel, aber sie postuliert mehr bedingte Wahrscheinlichkeiten beziehungsweise Zusammenhänge als zur Modellierung des Kontextes benötigt würden. Ein plausibles und inhaltlich korrektes Bayes'sches Netz muss entsprechend noch nicht sinnvoll nutzbar sein. Diese Schwäche basiert auf einer impliziten Annahme bei der Konstruktion Bayes'scher Netze: sowohl formal als auch intuitiv gilt für jedes Paar von Einflussfaktoren ohne gerichtete Verbindung implizit, dass diese Einflussfaktoren **unabhängig** voneinander sind. Im Rahmen der Erstellung des Bayes'schen Netzes können unterschiedlich strenge Anforderungen formuliert werden, die erfüllt sein müssen, bevor zwei Einflussfaktoren als unabhängig angesehen werden können. Beispielsweise kann gefordert werden, dass die Unabhängigkeit experimentell überprüft werden muss. Tatsächlich kann sogar im genannten Beispiel angenommen werden, dass eine Autowäsche vermutlich nicht an Tagen, an denen es regnet, stattfinden wird. Die Straße wird durch diese Verteilung auf mehrere Tage häufiger nass sein, als sie es wäre, wenn die beiden Einflussfaktoren völlig unabhängig voneinander wären. Der Versuch der Vermeidung solcher Fehler durch die Einfügung von zusätzlichen Zusammenhängen zwischen Einflussfaktoren führt häufig zu einer Überanpassung der Bayes'schen Netze.

Bayes-Klassifikatoren

Die Betrachtung von Assoziationslernen und Bayes'schen Netzen diene im Bisherigen der Darstellung des Einflusses der mathematischen Stochastik auf das maschinelle Lernen. Darüber hinaus lassen sich zwei der wichtigsten Klassifikatoren des maschinellen Lernens als Übergang von Assoziationslernen zu Bayes'schen Netzen verstehen. Diese beiden Algorithmen sind der **naive** und der **optimale Bayes-Klassifikator**.

Die diesen Klassifikatoren zugrunde liegende Beobachtung ist, dass das Hauptproblem bei der Manipulation und Einschätzung von Einflussfaktoren deren komplexe Abhängigkeiten voneinander darstellen. Wie bereits angedeutet wurde, kann das Wetter die Wahrscheinlichkeit einer Autowäsche beeinflussen oder ein Medikament kann in Kombination mit einem anderen Wirkstoff unerwünschte Nebenwirkungen entstehen lassen. Diese Abhängigkeit von Einflussfaktoren führt zu enorm komplexen Strukturen und erschwert das statistische Lernen. Der einfachste Ansatz zur Vereinfachung dieser Situation besteht darin, dass pauschal und mitunter wider besseres Wissen angenommen wird, dass die gemessenen Attribute der Trainingsinstanzen **statistisch unabhängig** sind. Zwei Größen sind statistisch unabhängig voneinander, wenn die bedingten Wahrscheinlichkeiten genau so groß sind wie die unbedingten Wahrscheinlichkeiten. Wenn etwa ein Studiengang die gleiche Geschlechteraufteilung aufweist wie die Gesamtbevölkerung, dann ist die Einschreibung in diesen Studiengang statistisch unabhängig von dem Geschlecht. Ein Gegenbeispiel wäre die Größenverteilung in einer professionellen Basketballmannschaft und in der Gesamtbevölkerung. Die Tätigkeit als Profibasketballer ist offenbar nicht unabhängig von der Körpergröße. Völlige statistische Unabhängigkeit liegt in der Realität fast nie vor, wird aber dennoch angenommen⁶¹. Mit Hilfe der aus dieser Annahme resultierenden Vereinfachungen bezüglich der bedingten Wahrscheinlichkeiten wird der Einsatz von neuen statistischen Autoadaptionsprozessen möglich. Das zentrale resultierende Verfahren zum Einsatz von

⁶¹ Natürlich können nur die Trainingsdaten als unabhängig angenommen werden, da sonst keinerlei Kanten im entstehenden Bayes'schen Netz enthalten wären, weil die Einflussfaktoren keine Einflüsse ausüben. Dennoch kann auch für die übrigen Größen eine etwas reduzierte Form von Unabhängigkeit angenommen werden: die **bedingte Unabhängigkeit**. Hierbei wird eine Größe als von genau einer Größe statistisch abhängig betrachtet, wie etwa im Falle einer Kausalität.

bedingten Wahrscheinlichkeiten zur Klassifikation von Eingabedaten wird aufgrund der notwendigen, aber kontrafaktischen Vereinfachung als naiver Bayes-Klassifikator bezeichnet. Naive Bayes-Klassifikatoren beruhen wie die Suche nach interessanten Assoziationsregeln darauf, die Anzahl der zu betrachtenden bedingten Wahrscheinlichkeiten so gering wie möglich zu halten, nutzen jedoch Bayes'sche Netze zur Darstellung ihrer Struktur.

Die Auswertung eines Bayes-Klassifikators entspricht den nachfolgenden Schritten **A** bis **E**. Hier ist festzuhalten, dass gezielt die Auswertung und nicht die Erstellung eines Bayes-Klassifikators dargestellt wird, da die Klassifikatoren als solche für das maschinelle Lernen von größerer Bedeutung sind als die Methode ihrer Erstellung.

- A** Die Attribute eines Eingabedatums werden als statistisch unabhängig betrachtet und einzeln ausgewertet.
- B** Ein Attribut des Eingabedatums wird ausgewählt und weiteren Schritt **B** wird als Bedingung angenommen, dass das Attribut denjenigen Wert aufweist, den es für das Eingangsdatum annimmt. Davon ausgehend wird jeweils berechnet, wie groß die bedingte Wahrscheinlichkeit ist, dass das Eingabedatum einer der im Problemkontext vorgegebenen Klassen angehört.

Dieser Schritt ist praktisch schnell verständlich. Angenommen Passanten, sollen bezüglich der Wahrscheinlichkeit eingeschätzt werden, dass sie Teil eines professionellen Basketballteams sind. Zu diesem Zweck werden verschiedene Daten erhoben, unter anderem die Körpergröße, die Sprunghöhe und die Größe der Hände⁶². Im Schritt **B** wird zuerst die Größe der Hände des Passanten als Attribut ausgewählt und mit derjenigen aller gemessenen Profi-Basketballer und der Restbevölkerung verglichen. Die beiden Häufigkeiten werden in ein Verhältnis gesetzt und bilden die Wahrscheinlichkeit Profisportler zu sein unter der Bedingung, Hände einer gewissen Größe zu besitzen.

- C** Der Schritt **B** wird für alle Attribute des Eingabedatums durchgeführt.
- D** Jede Wiederholung von Schritt **B** erzeugt für jede mögliche Klassenzuordnung eine Wahrscheinlichkeit. Die entstandenen attributsabhängigen Wahrscheinlichkeiten werden kombiniert⁶³.
- E** Die Klasse mit der größten kombinierten Wahrscheinlichkeit wird ausgewählt.

Die Nutzung von naiven Bayes-Klassifikatoren bildet neben Entscheidungsbäumen und KNN eines der oder sogar das im maschinellen Lernen am häufigsten eingesetzte Verfahren (Grieser et Fürnkranz 2006; Russell et al. 2007). Die resultierenden Klassifikatoren liefern häufig gute Ergebnisse, insbesondere solange die Einflussfaktoren in der Realität keinen zu starken Einfluss aufeinander ausüben. Varianten von Gewichtungen der Urteile einer Zusammenstellung naiver Bayes-Klassifikatoren bilden einige der effektivsten allgemein einsetzbaren Algorithmen im maschinellen Lernen, etwa bei Text-Klassifikationen im Rahmen derer Nachrichten bestimmten Themenfeldern zugeordnet werden sollen (Rennie 2001).

⁶² Alle drei Daten sind prinzipiell voneinander abhängig, können aber dennoch beim Einzelnen sehr stark abweichen, das bedeutet hier ist die Annahme einer statistischen Unabhängigkeit vielversprechend.

⁶³ Die Wahrscheinlichkeiten werden multipliziert und das maximale Produkt wird im nächsten Schritt ausgewählt.

Die Motivation der Annahme statistischer Unabhängigkeit der Attribute bestand darin, dass anderenfalls eine zu große Anzahl von bedingten Wahrscheinlichkeiten denkbar und deren Berechnung sehr aufwendig wäre. Die Vorgehensweise des naiven Bayes-Klassifikators kann abgesehen von Schritt **A** formal auch umgesetzt werden, ohne die vereinfachende Annahme der statistischen Unabhängigkeit zu machen, wenngleich die Schritte **B** bis **D** in diesem Fall mathematisch einiger weiterer Erklärung bedürfen. Diese Lösung wird als **optimaler Bayes-Klassifikator** bezeichnet und sie erzielt im Mittel die besten Ergebnisse aller maschinellen Lernverfahren. Dieses Verfahren ist zwar in der Praxis selten realisierbar, da es extrem aufwendig ist sämtliche notwendigen Wahrscheinlichkeiten zu bewerten, der optimale Bayes-Klassifikator kann jedoch als eine Kenngröße verwendet werden, um in kleinen Testfällen die Performanz anderer Lernverfahren einschätzen zu können.

2.3.7. Analytisches Lernen

Beim analytischen Lernen wird versucht mit Hilfe eines vorgegebenen Wissens aus Beobachtungen Hypothesen abzuleiten. Diese Hypothesen können dann dem vorgegebenen Wissen hinzugefügt werden, um die Wissensbasis für zukünftige Lernvorgänge zu erweitern.

(Spix 1998)

Analytisches Lernen kann in verschiedenen Ausprägungen auftreten. Einerseits kann **deduktives Lernen** umgesetzt werden, das auf der Manipulation bereits bestehenden Vorwissens beruht, auf Messwerte verzichtet und neue Aussagen aus bisherigen Aussagen folgert. Andererseits kann eine induktiv-deduktive Mischform umgesetzt werden, die aus Beispielen allgemeine Regeln extrahiert. Im Weiteren wird primär auf die induktive logische Programmierung als ein Beispiel für solch eine Mischform eingegangen. Zwar sind beide Ausprägungen des analytischen Lernens im maschinellen Lernen realisierbar, der Verzicht auf Messwerte beim deduktiven Lernen macht dieses jedoch zu einem Grenzfall.

Induktive logische Programmierung

Motivation

Das Ziel der **induktiven logischen Programmierung**, kurz **ILP**, besteht in der automatischen Generierung von Hintergrundwissen in Form von **Regeln**, die als logische Aussagen der Form „alle Menschen kreisen um die Sonne“ formuliert werden sollen. Während das Bayes'sche Lernen bedingte Wahrscheinlichkeiten als Mittel zur Struktursuche nutzt und Entscheidungsbäume als aussagenlogische Strukturen interpretiert werden können, manipuliert eine ILP Elemente der Prädikatenlogik, um in der Lage zu sein Regeln und damit Strukturvorschläge erstellen zu können⁶⁴. Das Problem, das durch den Einsatz von Prädikatenlogik gelöst wird, besteht darin, dass manche Zusammenhänge, die sich in den Trainingsdaten widerspiegeln, nicht ohne größere Umwege als Werte von Attributen beschreibbar sind, ein Beispiel ist der Zusammenhang „ist Mutter von“.

Die induktive logische Programmierung spielt wie das deduktive Lernen und die noch dargestellten Stützvektormethoden im zweiten Hauptteil keine zentrale Rolle. Aus diesem

⁶⁴ Die bei der ILP eingesetzten eingeschränkten Formen der Prädikatenlogik weisen eine höhere Darstellungskraft auf als Entscheidungsbäume, die auf Aussagenlogik basieren (Wikipedia Contributors 2012, Induktive logische Programmierung).

Grund wird sich die Darstellung darauf beschränken, herauszustellen, welche neuen Aspekte die induktive logische Programmierung der Diskussion des maschinellen Lernens beisteuern kann.

Einführungsbeispiel

Ein Beispiel für die Zusammenhänge, die in der ILP behandelt werden, ist der folgende Stammbaum.

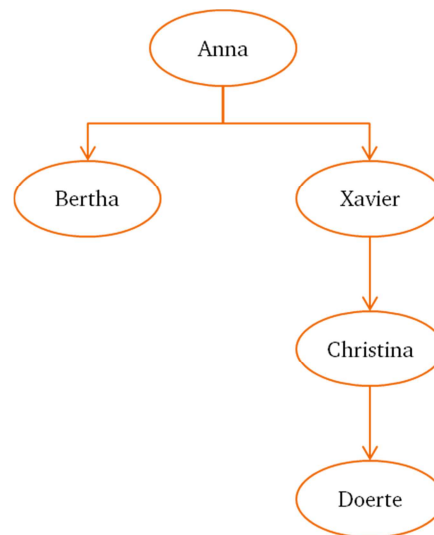


Abbildung 36: Stammbaum als Einführungsbeispiel zur ILP

Angenommen, hierzu wäre als Hintergrundwissen bekannt, welche der Personen wessen Kind ist und welche der Personen weiblich sind. Der Zusammenhang beziehungsweise das Konzept „ist Mutter von“ hingegen soll gefunden werden. Die Erstellung dieses Konzeptes setzt das Vorliegen von Trainingsdaten in Form von positiven und negativen Beispielen voraus. Sei entsprechend angenommen, dass die folgenden beiden Zusammenhänge als wahr bekannt sind:

- „Anna ist die Mutter von Bertha“
- „Christina ist die Mutter von Doerte“

Weiter seien die folgenden beiden Zusammenhänge als falsch bekannt:

- „Xavier ist die Mutter von Christina“
- „Anna ist die Mutter von Christina“

Die ILP wird jetzt versuchen, anhand der Beispiele eine logische Regel für das Konzept „A ist Mutter von B“ zu entwickeln. Eine solche Regel würde voraussichtlich aussagen „B ist Kind von A und A ist weiblich“.

Funktionsbeschreibung und Definition

Häufig entwickelt die ILP Regeln, die Trainingsdaten korrekt prognostizieren und verwirft dann die entsprechenden Trainingsdaten. Diese ILP arbeiten entsprechend in gewisser Weise gegensätzlich zum instanzbasierten Lernen, das die Trainingsdaten möglichst und langfristig intensiv nutzen will. Der resultierende Autoadaptionsprozess des ILP kann wie folgt zusammengefasst werden.

-
- A** Der gesuchte Zusammenhang wird mittels Trainingsdaten in Form von positiven und negativen Beispielen beschrieben.
 - B** Eine Aussagenmenge von Hintergrundwissen zu den Trainingsdaten wird etabliert.
 - C** Es wird eine Aussage gesucht, aus der sich mindestens ein positives Beispiel der Trainingsdaten herleiten lässt und aus der sich kein Negativbeispiel herleiten lässt⁶⁵.
 - D** Wenn im Schritt **C** eine Aussage gefunden wurde, wird sie zum Hintergrundwissen hinzugefügt, die Trainingsdaten werden um das Positivbeispiel reduziert und Schritt **C** wird wiederholt.
 - E** Wenn im Schritt **C** keine Aussage gefunden wurde oder keine Positivbeispiele mehr zu erklären sind, wird der Prozess abgebrochen und das erweiterte Hintergrundwissen wird als Strukturvorschlag ausgegeben.

Im Allgemeinen liefert diese Vorgehensweise eine Regel, aus der sich einige der positiven, jedoch kein negatives Beispiel herleiten lassen. Die Forderung, dass keine negativen Beispiele herleitbar sein dürfen, kann gelockert werden, indem nur gefordert wird, dass die Anzahl der positiven Beispiele deutlich größer sein muss als die der negativen Beispiele⁶⁶. Eine andere Modifikation des ILP ist es, zu fordern, dass das Hintergrundwissen vollständig im Rahmen des Autoadaptionsprozesses generiert werden muss.

Deduktives Lernen

Deduktives Lernen erweitert bereits bestehendes Vorwissen, indem neue Aussagen aus bereits bekannten Aussagen gefolgert werden, und verzichtet dabei auf die Adaption des Vorwissens auf Basis von Messwerten. Deduktives Lernen versucht zwar, so wie das übrige maschinelle Lernen, Strukturen zu finden, die zu vorliegenden Trainingsdaten passen, allerdings ist es dennoch ein Grenzfall maschinellen Lernens, da die Strukturvorschläge ausschließlich aus dem im Vorfeld gegebenen Vorwissen abgeleitet werden. Die Trainingsdaten werden nur herangezogen um zu prüfen, welche der hergeleiteten Strukturvorschläge im konkreten Kontext relevant sind⁶⁷.

Das deduktive Lernen ist aus interdisziplinärer Perspektive insofern interessant, als es den Grenzbereich zwischen maschinellem Lernen und nichtlernenden Algorithmen beleuchtet. Aus diesem Grund soll anhand zweier Varianten deduktiven Lernens ein Einblick in diesen Teil des maschinellen Lernens ermöglicht werden.

⁶⁵ Formal sind hierbei nur Aussagen zulässig, in denen nur Variablen als Argumente vorkommen, da sonst wiederum eine zu große – rein formal eine unendlich große – Anzahl von Aussagen denkbar ist und geprüft werden muss.

⁶⁶ Die Differenz der beiden Anzahlen wird als der **heuristische Wert** einer Regel bezeichnet.

⁶⁷ **Wissenslevel-Lernen** als Variante des deduktiven Lernens wird auch als **relevanzbasiertes Lernen** bezeichnet, allerdings ist die Einschätzung der Relevanz des Vorwissens allem deduktiven Lernen gemein – wenn auch in unterschiedlicher Ausprägung.

Erklärungsbasiertes Lernen

Das **erklärungsbasierte Lernen** sucht in einem gegebenen Vorwissen nach einer Erklärung für ein Trainingsdatum. Anschließend wird diese Erklärung verallgemeinert und als Strukturvorschlag festgehalten.

Ein Beispiel: 'Ein intelligenter Höhlenmensch brät seine erbeutete Eidechse an einem angespitzten Stock über einem Lagerfeuer, um sich seine Finger nicht am Feuer zu verbrennen. Seine weniger intelligenten Genossen, die zu diesem Zweck bislang ihre Finger benutzen, beobachten ihn dabei. Aus dieser Beobachtung und ihrem Hintergrundwissen können sie ableiten, daß man eine Eidechse rösten kann, ohne sich die Finger dabei zu verbrennen, indem man einen dünnen, spitzen Stock benutzt. Durch eine Generalisierung kommen sie zu dem Schluß, daß sich jedes Kleintier mit einem dünnen, langen, festen, spitzen Gegenstand gefahrlos über einem Feuer rösten läßt.'

(Spix 1998)

Die Erkenntnis, dass Kleintiere sich auf diese Weise rösten lassen, lässt sich formal auch ohne die Beobachtung aus dem Vorwissen der Höhlenmenschen ableiten. Die Beobachtung dient eher einer Inspiration als der Vorwegnahme eines Experimentes.

Wissenslevel-Lernen

Wissenslevel-Lernen sucht formal nicht nach einer Erklärung für Trainingsdaten, sondern nach Anwendungskontexten für Vorwissen. Diese Variante deduktiven Lernens nutzt Trainingsdaten, um abstraktes Wissen mit Hilfe von Trainingsdaten zu konkretisieren.

Ein Beispiel: 'Eine Reisende kommt erstmals nach Brasilien und trifft ihren ersten Brasilianer. Sie hört ihn Portugiesisch sprechen und erkennt, daß sein Name Fernando ist. Aufgrund ihres Vorwissens, daß innerhalb eines Landes die meisten Bewohner eine Sprache sprechen, folgert sie, daß Brasilianer Portugiesisch sprechen, jedoch folgert sie nicht, daß alle Brasilianer Fernando heißen, da Namensgleichheit nicht eine allgemeingültige Eigenschaft der Bewohner eines Landes ist.'

(Spix 1998)

Die Menschen in Brasilien hätten ausgehend vom Vorwissen der Reisenden auch jede andere Sprache sprechen können, sie wusste zunächst nur, dass die ihr unbekannte Muttersprache für die meisten Einwohner identisch ist. Die Aussage, dass in Brasilien alle Menschen Portugiesisch sprechen, beinhaltet somit sowohl mehr als das Vorwissen als auch mehr als eine Beschreibung der Beobachtung, das heißt die Trainingsdaten wurden in gewisser Hinsicht doch als Messwert genutzt.

2.3.8. Stützvektormethoden

Motivation

Analog zu statistischem Lernen basieren auch **Stützvektormethoden** nicht auf der Idee eine Form von Selbstorganisation zu mathematisieren, sondern auf der Implementierung

mathematischer Optimierungsverfahren in Kontexten, die von Eingabedaten abhängen. Die grundsätzlichen Ideen hinter den zugrunde liegenden Optimierungsverfahren wiederum entstammen der **mathematischen Optimierung** und nicht der Informatik und sind für eine interdisziplinäre Betrachtung des maschinellen Lernens entsprechend nicht zentral. Im Gegensatz zu statistischem Lernen wird bei Stützvektormethoden darüber hinaus nicht auf dem Umgang mit einem auch außerhalb der Mathematik bedeutsamen Konzept – wie es die Wahrscheinlichkeit war – aufgebaut. Stützvektormethoden selbst sind dennoch in der Praxis von einiger Bedeutung, vor allem wenn klare Zielvorgaben formulierbar sind. Diese Situation kann auch durchaus eintreten, nachdem bereits ein anderer Ansatz des maschinellen Lernens eingesetzt wurde und die Performanz des Strukturvorschlages erhöht werden soll.

Einführungsbeispiel

Stützvektormethoden basieren auf der Trennung von Trainingsdaten. Eine Veranschaulichung für solch eine Trennung ist die nachfolgende Auswertung einer fiktiven Umfrage zur Entgeltgleichheit.

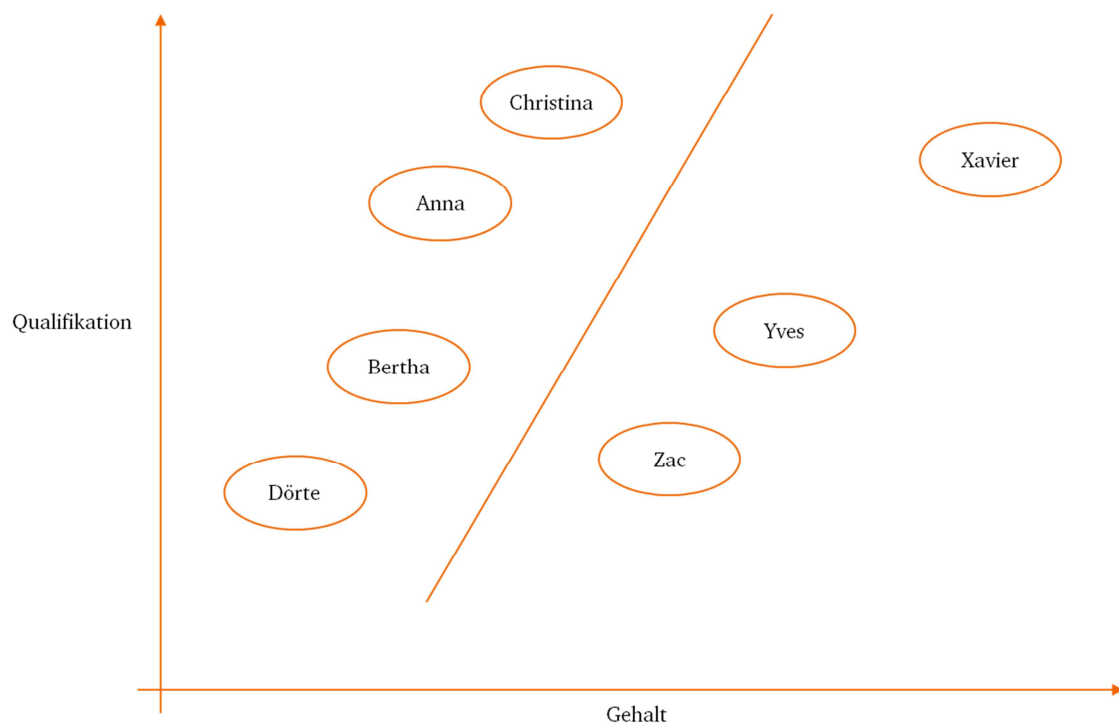


Abbildung 37: Fiktive Grafik zum Entgelt von Angestellten

Hier können die Trainingsdaten ohne Schwierigkeiten getrennt werden und stehen somit bereit für eine weitere Analyse durch die Nutzer.

Funktionsbeschreibung

Stützvektormethoden suchen trennende Geraden, Ebenen oder höherdimensionale Ebenen, die als **Hyperebenen** bezeichnet werden. Die Nutzung von Hyperebenen setzt eine Codierung voraus, bei der die Trainingsdaten als Punkte in mehrdimensionalen Räumen identifiziert werden. Die Trennung mittels Hyperebenen ist insofern verwandt mit dem instanzbasierten Lernen, als auch Stützvektormethoden einen Abstandsbegriff voraussetzen und den Suchraum in Teilräume aufspalten, um Eingabedaten klassifizieren zu können. Das Ziel der Trennung liegt häufig darin, eine Trennung mit einer möglichst großen neutralen Zone zwischen den

getrennten Trainingsdaten zu identifizieren. Die Zone II in der folgenden Visualisierung wird gegenüber der Zone I bevorzugt.



Abbildung 38: Lineare Trennung mit kleiner (I) und großer (II) neutraler Zone

Die dem Rand am nächsten liegenden Trainingsdaten werden als **Stützvektoren** bezeichnet und sobald die Stützvektoren identifiziert wurden, spielen die übrigen Trainingsdaten für die Stützvektormethoden meist keine Rolle mehr – insofern sind die Unterschiede zum instanzbasierten Lernen deutlich erkennbar.

Stärken und Schwächen

Stützvektormethoden sind in der Praxis insbesondere für Aufgaben in der Bilderkennung, wie etwa Handschrifterkennung, sehr gut geeignet. Im Praxisbeispiel zu künstlichen neuronalen Netzen wurde dieser Anwendungsfall bereits betrachtet, tatsächlich wurden jedoch sowohl KNN als auch Stützvektormethoden eingesetzt. Die Fehlerrate der betrachteten KNN konnte, wie bereits erwähnt, von 1,6% auf 0,7% verbessert werden, während die Fehlerrate bei Stützvektormethoden im gleichen Kontext ohne Berücksichtigung von Vorwissen zunächst bei 1,1% lag und auf 0,56% gesenkt werden konnte (Russell 2007, S. 914ff). Auch im Allgemeinen können Stützvektormethoden besonders in Kontexten eingesetzt werden, in denen die Trainingsinstanzen eine Vielzahl von Attributen aufweisen, die sich als Zahlenwert codieren und somit ohne Weiteres als Punkte im Raum interpretieren lassen – wie etwa die Farbe von Bildpunkten bei der Bilderkennung. Stützvektormethoden sind damit generell in ähnlichen Kontexten wie KNN einsetzbar. Ein Vorteil der Stützvektormethoden liegt darin, dass die Codierung von Daten als Punkte im Raum die Eingabereihenfolge der Attribute der Daten irrelevant werden lässt. Das wiederum impliziert, dass Strukturen, die sich in der Eingabereihenfolge verbergen, nicht ohne einen Umweg entdeckt werden können.

Eine weitere Schwäche der Stützvektormethoden besteht darin, dass eine zu große Anzahl irrelevanter Attribute nur schwer gehandhabt werden kann. Zwar können die Attribute formal berücksichtigt werden, allerdings ist in der Grundidee der Stützvektormethoden nicht mitgedacht, dass unterschiedliche Gewichtungen für die Abstände zwischen wichtigen und unwichtigen Attributen sinnvoll sein können. Das Auftreten irrelevanter Attribute führt damit bei Stützvektormethoden zu einer Überanpassung, da jede Abweichung und jeder Abstand als gleich bedeutsam beurteilt wird.

2.4. Charakteristik des maschinellen Lernens

Der Übergang von der Darstellung der Perspektive der Informatik zur Verortung des maschinellen Lernens durch die Technikphilosophie kann entlang der Frage vollzogen werden, ob die Autoadaptionsprozesse des maschinellen Lernens einer Suche, Optimierung, Klassifikation oder keiner der drei Charakterisierungen entsprechen. Die im Vorherigen vollzogene Aufspaltung des maschinellen Lernens in Teilgebiete, die durch ihre zugrunde liegenden Ideen unterschieden wurden, bietet interdisziplinären Betrachtungen eine stabile Grundlage. Selbstverständlich wurden und werden in der Informatik über die genannten Varianten maschinellen Lernens hinaus noch eine Vielzahl von weiteren Ansätzen für maschinelles Lernen entwickelt. Unabhängig davon, wie erfolgreich oder prominent diese Ansätze relativ zu den vorgestellten Varianten maschinellen Lernens zu bewerten sind, liegt der Fokus der jeweiligen Entwicklung nur sehr selten auf der Entwicklung einer systematisch neuen Art maschinell zu lernen. Stattdessen werden Kombinationen verschiedener Lernstrategien, Weiterentwicklungen bestehender Algorithmen und insbesondere Anpassungen von MLA an konkrete Kontexte realisiert. Solche Maßnahmen haben – bisher – ebenso wie Mischformen der bereits beschriebenen Verhaltensweisen keine systematisch neuartigen Verhaltensweisen bei MLA entstehen lassen. In Hinblick auf die übergreifende Frage nach der Selbstorganisation und der Veränderung von Technik sind solche Anpassungen dementsprechend ebenso wenig von zentraler Bedeutung wie bezüglich der Diskussion der Fragen, wie ein Autoadaptionsprozess sich charakterisieren lässt. Die bisherige Darstellung der wesentlichen Konzepte zur Erstellung von Autoadaptionsprozessen stellt die Grundlage für die Einsicht dar, dass eine pauschale Antwort auf die Fragen nach der Prozesscharakteristik des maschinellen Lernens als Gesamtgebiet nicht zielführend ist. Eine Antwort auf diese Frage ist hochgradig von dem betrachteten Teilgebiet des maschinellen Lernens abhängig.

Auch wenn eine pauschale Einordnung des maschinellen Lernens nicht sinnvoll möglich ist, können sehr wohl einige der in der Diskussion gebräuchlichen Begriffe als aus technikphilosophischer Sicht ungeeignet bestimmt werden. Die Identifikation von Autoadaptionsprozessen mit der nachträglich zugeschriebenen Funktion der aus den Autoadaptionsprozessen resultierenden Strukturvorschläge wie bei der Beschreibung eines MLA als **Klassifikator** ist ein solcher Fall. Die Interpretation eines Autoadaptionsprozesses als eine **Suche** ist zwar prinzipiell möglich, bietet aus technikphilosophischer Sicht jedoch ebenfalls keinen großen Mehrwert, denn die Durchführung einer Suche impliziert die Vorgabe eines zu Suchenden und gewisser Bewertungskriterien. Eine Suche unterscheidet sich somit nur insofern von einer **Optimierung**, als dass die Rede von einer Suche eine weniger systematische Vorgehensweise annimmt. Auch wenn die Unterscheidung zwischen einer Optimierung und einer Suche nicht unmittelbar hilfreich ist, vermittelt die Tatsache, dass im maschinellen Lernen der Begriff der Suche verwendet wird, einen Eindruck von der Denkweise der Informatik über das maschinelle Lernen. Einige Ansätze des maschinellen Lernens formulieren durchaus den Anspruch sich von einer reinen Optimierung abzusetzen. Eine mögliche technikphilosophische Entsprechung dieser Abgrenzungsversuche und der resultierenden Formen von Artefakten wird im Folgenden entworfen.

3. Zweiter Hauptteil: Der Blick der Technikphilosophie auf maschinelles Lernen

Im Folgenden wird untersucht, welche Teile des maschinellen Lernens aus technikphilosophischer Sicht gegenüber klassischen Algorithmen eine veränderte Form von Technik darstellen und wie sich diese Veränderungen begrifflich erfassen lassen. Zu diesem Zweck werden im Weiteren nach einer Problemeröffnung und einer Begriffsklärung einige Perspektiven der Technikphilosophie daraufhin untersucht, welche Aspekte maschinell lernender Artefakte sie jeweils angemessen beschreiben können. Die Diskussion zielt dabei auf die Erstellung eines Beschreibungsvorschlags ab, der darstellt, was genau an maschinell Lernen als systematisch neu betrachtet werden kann. Die entsprechende Teilgruppe der MLA wird im Weiteren identifiziert, beschrieben und als neugieriges maschinelles Lernen beziehungsweise als Welttechnik bezeichnet werden. Die Vorgehensweise wird einem Ausprobieren entsprechen – das heißt einem konstruktiven Ablaufen von begrifflichen Sackgassen – und somit eine Annäherungsbewegung darstellen, bei der jeder Enttäuschungsschritt einen Gewinn mit sich bringt und die Problemstellung präzisiert. Diese Ökonomie der Gewinne und Verluste im Versuchsaufbau zielt darauf ab, das Problem durch methodische Tests der Passung von Perspektiven einzukreisen. Hierfür wird der erste Hauptteil als Grundlage benötigt, denn die Phänomene des maschinellen Lernens sollen zwar mit einer anderen Sprache konfrontiert werden, diese soll aber noch so nahe an den Inhalten sein, dass die Analyse ihren interdisziplinären Charakter behält und nicht zu einer geisteswissenschaftlichen Außenansicht wird. Nicht zuletzt sollen auch Diskutanten aus der Informatik aus der Analyse einen Mehrwert gewinnen können.

Als Einleitung in die Diskussion wird deren Methodik zunächst exemplarisch am Beispiel der künstlichen neuronalen Netze verdeutlicht. Die bisherige Darstellung von KNN hatte ihren Schwerpunkt in der Abgrenzung zu anderen Lernstrategien und der Einführung in das maschinelle Lernen. Im Folgenden werden daher zunächst die Eigenschaften der KNN in Hinblick auf die Frage nach der Neuartigkeit maschinellen Lernens rekapituliert.

3.1. Präzisierung der Diskussion am Beispiel von KNN

Maschinelles Lernen umfasst verschiedene Ansätze, Artefakte zu konstruieren, die in unbekannten oder chaotischen Umgebungen Strukturen suchen. Mit den künstlichen neuronalen Netzen wird zunächst ein solcher Ansatz analysiert und die nachfolgende Diskussion vorbereitet, die sich mit verschiedenen gebräuchlichen Perspektiven beschäftigen wird, autoadaptive Artefakte begrifflich zu fassen. KNN eignen sich sehr gut als exemplarische Lernstrategie, weil sie sich durch ein besonders geringes Maß an Erwartbarkeit auszeichnen. Tatsächlich ist die Erwartbarkeit so schwach ausgeprägt, dass KNN in industriellen Kontexten fast nicht in ihrer Grundform eingesetzt werden. Die Anwendungen finden sich eher in der Umsetzung von Brettspielen und vergleichbaren Kontexten⁶⁸. Die KNN wurden in Abschnitt 2.3.4 bereits vorgestellt, die Begriffsbildung soll daher nur kurz rekapituliert werden. Mit den Neuronen wurde ein Konstruktionselement biologischer Gehirne aus seinem extrem komplexen und bisher noch nicht verstandenen Kontext gelöst, formalisiert und mathematisch für die Informatik nutzbar gemacht. Strukturen der entstandenen künstlichen Neuronen werden künstliche neuronale Netze genannt und obwohl der Anspruch künstlicher neuronaler Netze nicht darin liegt, biologische Gehirne nachzubilden, teilen KNN einige ihrer Stärken und Schwächen.

[...] Damit wird verständlich, wie das Lernen nicht funktioniert, nämlich indem Nervennetze (wie Computer) Informationen speichern und Wissen akkumulieren würden. Vielmehr bildet das Gehirn Informationen, und zwar auf der Grundlage von dynamischen, sich selbstorganisierenden Netzwerken und nur unter dem Einfluss eigener Aktivitäten (Prinzip der aktivitätsgesteuerten Anpassung).

(Teuchert-Noodt 2011)

3.1.1. Vorstrukturierte künstliche neuronale Netze

Die Netze natürlicher Neuronen in biologischen Gehirnen sind zwingend sehr stark und systematisch vorstrukturiert. **KNN** lassen sich auf Basis theoretischen Vorwissens ebenfalls vorstrukturieren, so dass sie nicht in einem zufälligen oder gleichförmigen Zustand den Lernbeziehungsweise Autoadaptionsprozess beginnen – allerdings ist die Vorstrukturierung bei KNN optional.

Wenn im Vorfeld des Einsatzes eines MLA ein klares Ziel formulierbar und entsprechendes theoretisches Vorwissen vorhanden ist, wird fast immer auch eine Vorstrukturierung vorgenommen, da das KNN die aus dem Vorwissen resultierenden Konzepte aller Voraussicht nach sowieso nachbilden muss. Vorstrukturierte Netze lösen somit vorformulierte Probleme und erreichen ihr Ziel meist auf eine in bestimmter Hinsicht als optimal klassifizierbare Weise. Der Maßstab, nach dem die Qualität der Lösung eingestuft wird, ist dabei ebenfalls im Vorhinein gegeben und kann entweder in die Struktur des Netzes integriert werden oder dem Autoadaptionsprozess nachgelagert sein. Ein Beispiel für solch eine Problemlösung ist die Suche nach Gewinnstrategien in Brettspielen wie Schach oder Backgammon. Eine

⁶⁸ Brettspiele bieten eine gut beschreibbare und kontrollierbare Umgebung für den Einsatz von maschinell lernenden Artefakten und eignen sich daher besonders für die Analyse von unerwartetem Verhalten und den Nachweis von Verbesserungen in der Performanz der zugrunde liegenden Algorithmen. Gleichzeitig ist unerwartetes Verhalten beim Einsatz als Gegenspieler eines Menschen durchaus erwünscht, wenn die optimalen Vorgehensweisen im Vorfeld noch nicht bekannt sind.

Vorstrukturierung könnte hier etwa dafür sorgen, dass Bauern nur in Damen umgewandelt werden oder dass die Wahrscheinlichkeiten für das Auftreten gewisser Augenzahlen beim Würfeln präzise vorgegeben werden. Beim Schach entscheiden Menschen über ihren **Halbzug**⁶⁹ vor allem, indem sie mögliche zukünftige Stellungen geistig durchspielen. Der Strukturvorschlag eines KNN entspricht nun genau einer solchen potenziellen Sequenz von jeweils optimalen Halbzügen. Ein menschlicher Spieler kann dementsprechend sofort nach Abschluss des Autoadaptionsprozesses dessen Ergebnisse in seiner eigenen Strategie berücksichtigen. Die Vorgehensweise des KNN muss nicht mehr weiter verstanden werden und der Autoadaptionsprozess kann als eine **Black Box**⁷⁰ zur Erstellung von Siegstrategien betrachtet werden. Die Nutzbarkeit dieser Black Box ist in der Praxis durch einen belastbaren Fundus an theoretischem Wissen sichergestellt, das in die Vorstrukturierung der Black Box investiert wurde. Auch wenn die Entscheidungsfindung eines KNN nicht nachvollziehbar ist, kann ein Nutzer mittels eigener Reflektionen über das bloße Befolgen der Zugvorschläge hinaus seine Spielweise verbessern. So hat sich etwa die Spielweise von Großmeistern – auch in deren eigener Wahrnehmung – durch die Nutzung von maschinellem Lernen deutlich verändert.

It's very important to take into consideration that computers have changed chess. The drawback of them is that many games begin only with the 35-th move or even later. So you are immediately involved in endgame and there is no live game.

(Doggers im Interview mit Spassky am 19.04.2009)

Bei Backgammon ist eine solche Verbesserung der eigenen Spielstärke durch Nutzung einer auf ein Ziel ausgerichteten Black Box bereits deutlich schwieriger. Im Kontext des Backgammons entstand durch den Einsatz von künstlichen neuronalen Netzen die Situation, dass die KNN erfolgreiche Züge empfehlen konnten, aber nicht erklärt werden konnte, was diese Züge erfolgreich macht.

It's not like people think, that you press the button and that the computer tells you what are the best moves and you go and play them. On such a level it's different. You go much deeper than the theory says. You have to look for where your opponent may try to surprise you so you basically have to recheck all the theory of the opening, learn and then recheck everything.

(Doggers im Interview mit Gelfand am 06.06.2012)

Insgesamt weist die Verwendung vorstrukturierter, eine optimale Lösung suchender KNN die Struktur **instrumentellen Handelns** auf. Der **innere Zweck**⁷¹ entsteht daraus, dass der menschliche Spieler seine Spielstärke im Schach oder Backgammon verbessern will. Das

⁶⁹ Als Halbzug wird im Schach die Bewegung einer einzelnen Figur bezeichnet. Zwei Halbzüge ergeben einen Zug, der entsprechend die Bewegung einer weißen und einer schwarzen Figur umfasst.

⁷⁰ Ein System dessen innere Abläufe nicht betrachtet werden – oder werden können – und das nur in Hinsicht auf seine Reaktionen auf Eingaben hin untersucht wird.

⁷¹ Details zu inneren und äußeren Mitteln und Zwecken finden sich bei Hubig (Hubig 2007, S. 232f; Hubig 2008 II).

korrespondierende innere Mittel besteht im Lernen aus der Betrachtung der Spielstrategie eines Schach- beziehungsweise Backgammonprogramms. Der äußere Zweck besteht in der Bereitstellung von Strukturvorschlägen, die in jeder Situation einen optimalen Zug vorschlagen, und das äußere Mittel ist die Erstellung solcher Strukturvorschläge mittels der Erstellung eines KNN unter der Nutzung von menschlichem Vorwissen. Es verbleibt die Aufgabe, die Differenz der beiden Zwecke zu betrachten und zu analysieren, welche Eigenschaften des KNN als äußerem Mittel auf welche Weise für dessen spielerische Überlegenheit verantwortlich sind. Eine solche Analyse ist jedoch häufig nicht oder nur mit großem Aufwand möglich.

Today, chess programs have become so good that even grandmasters sometimes struggle to understand the logic behind some of their moves.

(Rogoff 2010)

3.1.2. Prinzipielle Intransparenz von KNN

Im Rahmen des Autoadaptionprozesses modifiziert sich ein künstliches neuronales Netz so lange, bis es auf die Trainingsdaten auf eine erwünschte Art und Weise reagiert. Oft genügt es für den späteren Einsatz des Artefaktes jedoch noch nicht, zu wissen, dass dieser äußere Zweck erreicht wurde. Ein KNN reagiert möglicherweise nur genau im Zusammenhang mit den trainierten Daten wie gewünscht. Zur Vermeidung solch einer Überanpassung muss analysiert werden, warum KNN auf eine bestimmte Art und Weise reagieren. Diese Forderung ist jedoch nicht ohne Weiteres erfüllbar.

Any automatically trained net with more than a few dozen neurons is virtually impossible to analyze and understand. One can't tell if a net has memorized inputs, or is 'cheating' in some other way.

(Fraser 2003)

Diese Aussage Frasers zur Analysierbarkeit entstammt einer Diskussion der folgenden, fiktiven Illustration zum Problem der Überanpassung von KNN. Frasers Beispiel liest sich wie folgt (Fraser 2003): Ein Verteidigungsministerium gibt ein MLA in Auftrag, das es ermöglichen soll, getarnte Panzer mittels einer Bilderkennung zu identifizieren. Die Trainingsdaten bestehen aus 100 Fotos von im Unterholz verstecktem Kriegsgerät und 100 Fotos von friedlichen Waldstücken. Nachdem das KNN mit der Hälfte der Daten trainiert wurde und zuverlässig Panzer als solche klassifizieren kann, wird es mit der anderen Hälfte der Daten auf seine Zuverlässigkeit getestet und gibt ebenfalls die erwünschten Antworten. Das heißt, es scheint abgesichert, dass das KNN nicht lediglich die Bilder und die jeweils gewünschten Antworten zusammen abgespeichert hat. Anschließend externe Tests der Fähigkeit zur Suche nach Panzern resultieren allerdings dennoch in einem völlig zufälligen Ergebnis. Der Grund ist, dass für die Trainingsdaten alle Bilder von Panzern bei bedecktem Himmel angefertigt wurden und alle übrigen Bilder bei Sonnenschein entstanden. Das KNN hat in Konsequenz nur die Fähigkeit entwickelt, zuverlässig die Farbe des Himmels erkennen. Das Wesentliche an dieser Illustration der Überanpassung von KNN ist, dass oft nicht geprüft werden kann,

was genau ein KNN realisiert⁷² – wenn sich der entstandene Strukturvorschlag überhaupt als ein Konzept im menschlichen Sinne ausdrücken lässt. Meist kann nur der formale Strukturvorschlag – das nach Abschluss des Lernvorgangs fixierte Netz – untersucht werden und nur über dieses Resultat kann der heuristische Versuch einer Erklärung für die Reaktionen des Netzes unternommen werden⁷³. Entsprechend werden umfangreiches heuristisches Vorwissen und detaillierte Kenntnisse der theoretischen Informatik und Mathematik benötigt, wenn die Nutzbarkeit der Strukturvorschläge eines KNN über die Trainingsdaten hinaus abgesichert werden soll. Insgesamt können künstliche neuronale Netze eine Vielzahl von Kausalgesetzen abbilden und Strukturen in großen Datenmengen finden, etwa optimale Zugsequenzen in der riesigen Menge der im Schach zulässigen Halbzüge. Gleichzeitig ist jedoch eine Steuerung des zugrunde liegenden Autoadaptionsprozesses nur über die Vorauswahl der einzulesenden Sensordaten⁷⁴ oder über eine starke Vorstrukturierung möglich und die gezielte Erstellung von Strukturvorschlägen auf eine spezielle, nachvollziehbare Weise erfordert nutzerseitig ein großes Vorwissen. Spezifisch für künstliche neuronale Netze ist jedoch, dass sowohl die Vorauswahl der Sensordaten als auch die Vorstrukturierung unterlassen werden können und die MLA häufig dennoch einen auf unbekannte Weise systematischen Strukturvorschlag erstellen können. Diese MLA können demnach nicht nur für die Lösung von theoretisch durchdrungenen Optimierungsproblemen eingesetzt werden, sondern sie eröffnen die Möglichkeit zur Entdeckung von interessanten Strukturen beziehungsweise unbekannten Konzepten und können dementsprechend für die Erschließung eines neuen Raums technischen Handelns eingesetzt werden.

3.1.3. Optionale Zwecklosigkeit der Struktursuche

Der Begriff des **Zwecks** wird im Folgenden als die erste der eingangs genannten begrifflichen Sackgassen identifiziert, deren Gewinn in einem besseren Verständnis der Besonderheit maschinellen Lernens besteht. Konkret besteht der Gewinn in der Feststellung, dass es möglich ist, gezielt Artefakte zu konstruieren und einzusetzen, die eine prinzipielle Intransparenz aufweisen. Manchmal soll beim Einsatz von KNN gerade nichts im Vorhinein Spezifizierbares gefunden oder optimiert werden. Dies gilt insbesondere bei der Analyse riesiger Datenbanken. Ein nicht vorstrukturiertes KNN ist in der Lage, eine Datenbank auf eine Art und Weise zu kategorisieren, die zunächst keine analytische Begründung hat und die keiner vom Nutzer vorgegebenen Idee entspricht. Ein Nutzen kann aus solch einer Kategorisierung gewonnen werden, indem die Kategorien analysiert und interpretiert werden. Eine solche Interpretation wiederum kann beispielsweise darin bestehen, Kategorien als Produktcluster häufig zusammen verkaufter Produkte zu verstehen und der daraus resultierende Nutzen schließlich könnte eine Kaufempfehlung für die Kunden sein. Alternativ können die Kategorien Zeiträume abbilden, in denen bestimmte Produkte vermehrt gekauft werden, wodurch Einkaufsvolumina gesteuert werden können. Das folgende Szenario ist ein sehr bekanntes – fiktives – Beispiel für das Auftreten einer unvorhergesehenen Struktur bei der Aufbereitung riesiger Datenbanken.

⁷² In Hinblick auf noch unbekannte, zukünftige Eingaben.

⁷³ In Hinblick auf bereits erfolgte, vergangene Eingaben.

⁷⁴ Die Sensordaten können beispielsweise so ausgewählt werden, dass sie eine speziell zu erlernende Fertigkeit gut wiedergeben.

For example, by scanning each sale into a data warehouse, grocery stores have determined that men in their 20s who purchase beer on Fridays after work are also likely to buy a pack of diapers. Thus, a display of Pampers or another brand might be set up in the beer aisle, or merchants will put one (but not both) of the products on sale on Friday evenings.

(Fisk 2006)

Zur Suche nach Strukturen, wie den sich gegenseitig bedingenden Wahrscheinlichkeiten für den Kauf von Windeln und Bier, würde eher ein statistisch lernendes MLA zu Einsatz kommen als ein KNN. Dies ändert sich, wenn im Vorfeld nicht klar ist, dass bedingte Wahrscheinlichkeiten gesucht werden. Die Stärke und Besonderheit von KNN liegen darin, dass es möglich ist, sie im Rahmen des Autoadaptionsprozesses unbeeinflusst Strukturen ausbilden zu lassen. Das KNN beantwortet nicht eine konkrete Frage nach der interessantesten bedingten Wahrscheinlichkeit, sondern sucht nach unvorhergesehenen Strukturen in den Eingabedaten, wie es beispielsweise die adaptive Resonanztheorie realisiert. Ein entsprechender Autoadaptionsprozess kann mit zufälligen und selbstveränderlichen Anfangswerten gestartet werden und nicht nur die Eigenschaften der resultierenden Kategorien können komplett unvorhersehbar sein, sondern auch deren Anzahl muss sich nicht im Vorhinein abschätzen lassen. Wie bereits beschrieben, entsteht mitunter nur eine einzige Kategorie, die alle Daten enthält, oder es wird jedes Datum einzeln betrachtet. Weiter ist, selbst wenn die Anzahl handhabbar ist, nicht sicher, dass die Kategorisierung sich nach menschlichem Ermessen von einer rein zufälligen Einteilung unterscheidet. Die Hoffnung beziehungsweise der Wunsch dabei ist, dass der Nutzer durch die entstandenen Kategorien kreativ angeregt wird und spontan eine Möglichkeit erkennt, wie er den vorgestellten Strukturvorschlag sinnvoll einsetzen kann. Nutzer benötigen auch in Fällen, in denen ohne Vorstrukturierung entstandene Strukturvorschläge interpretiert werden sollen, theoretisches und heuristisches Vorwissen, um das Resultat des Lernvorgangs überhaupt auf irgendeine Art und Weise verwenden zu können – allerdings wird in solchen Fällen nur Vorwissen über den Anwendungskontext benötigt und nicht über theoretische Informatik. Unabhängig davon kann in diesen Fällen nicht mehr von der Suche einer nützlichen Struktur als einem Zweck beim Einsatz eines auf künstlichen neuronalen Netzen basierenden MLA gesprochen werden.

Sachverhalte sind nur soweit Zwecke, als ihre Herbeiführbarkeit durch mögliche Mittel unterstellt wird (sonst handelt es sich um Wünsche oder Visionen).

(Hubig 2007, S. 231)

Eine im ersten Teil dargestellte Schwäche des Einsatzes von KNN liegt darin, dass fast nie erkennbar ist, auf welche Art und Weise diese MLA Strukturvorschläge erstellen und im Nachhinein fast immer nur heuristisch überprüft werden kann, was genau die Strukturvorschläge abbilden. Demgegenüber steht jedoch die im Weiteren zentrale Stärke: durch die fehlende Notwendigkeit einer Vorstrukturierung von künstlichen neuronalen Netzen kann der Autoadaptionsprozess ohne explizites Ziel initiiert werden – lediglich auf Basis der Hoffnung, dass die zu erstellenden und im Vorfeld nicht festgelegten oder bekannten Strukturvorschläge sich auf eine nützliche Art werden interpretieren lassen. Eine solche Hoffnung ist jedoch ein Wunsch und kein Ziel oder Zweck, da nicht begründet

angenommen werden kann, dass eine nutzbringende Interpretation prinzipiell möglich sein wird oder gar wie eine solche Interpretation aussehen könnte. Die weitere Diskussion soll genau diesen **ziellosen** Einsatz von MLA und insbesondere von nicht vorstrukturierten künstlichen neuronalen Netzen betrachten. Dieser Einsatz betrifft in erster Linie Kontexte, in denen eine Vorstrukturierung entweder nicht möglich oder nicht gewollt ist. Diese Fälle bilden einen großen Teil der Arbeit mit nicht vorstrukturierten KNN, da für erschlossener Kontexte häufig andere, angepasste Ansätze des maschinellen Lernens besser geeignet sind, wie etwa statistisches Lernen im obigen Beispiel. Hieraus ergibt sich auch die Rede von einer *optionalen* Zwecklosigkeit, denn es erscheint zwar prinzipiell möglich die Struktursuche eines MLA ohne Ziel zu denken, aber solch eine **Suche** ist keineswegs der Normalfall im maschinellen Lernen. Nur ein Teil der im ersten Hauptteil dargestellten Lernstrategien hat diesbezüglich ein erhöhtes Potenzial und das statistische Lernen liegt nicht in diesem Teil der Lernstrategien.

Die Motivation zum ziellosen Einsatz von nicht vorstrukturierten KNN kann beziehungsweise sollte nicht die Struktur des Strukturvorschlages selbst betreffen, sondern kann sich nur auf die Möglichkeiten zum Umgang mit dem Strukturvorschlag beziehen. Wenn ein Schachprogramm trainiert wurde und der entstandene Strukturvorschlag die Eigenschaft hat, in jeder Stellung einen starken Zug empfehlen zu können, wird eine Analyse des MLA eher den Umgang mit den vorgeschlagenen Zugsequenzen aufarbeiten als sich mit den Gewichten der Verbindungen zwischen den Knoten des resultierenden KNN zu beschäftigen⁷⁵. Bezüglich des Autoadaptionsprozesses und des Strukturvorschlages können dementsprechend auch keine Soll-Ist-Vergleiche angestellt werden, denn beide entwickeln sich im Wesentlichen unerwartet. Nutzer können bei einem zwecklosen Einsatz von Netzen den Autoadaptionsvorgang oder ihr eigenes Verhalten entsprechend nicht auf Basis der Nutzungs-Wünsche reflektieren. Auch der Versuch einer Analyse der Autoadaption des MLA durch die Beobachtung der Reaktion auf Sensordaten ist schwierig. Zwar nehmen nicht vorstrukturierte KNN Sensordaten auf und nutzen diese als Anlass zur Autoadaption gemäß eines statischen oder wiederum adaptiven Algorithmus, allerdings erfolgt die Autoadaption nicht im Hinblick auf eine Zielfunktion, da der Algorithmus gerade kein Ziel verfolgt⁷⁶. Das bedeutet, eine Bewertung der beobachtbaren Reaktion eines MLA auf ein Trainingsdatum ist nicht ohne Weiteres möglich. Ohne die Steuerung durch eine Zielfunktion werden Sensordaten nicht als Störungen oder Gefahren interpretiert, die systematisch umgangen werden müssen, sondern entsprechen lediglich einer neutralen Messung. Das heißt, nicht vorstrukturierte KNN **regeln** oder **steuern** sich im Rahmen des Autoadaptionsprozesses nicht selbst und ein Nutzer kann entsprechend die Übergabe eines Trainingsdatums nicht als Steuerungsimpuls nutzen, um auf diesem Weg den Autoadaptionsprozess zu reflektieren. Genauso problematisch ist schließlich der Versuch, den Einsatz nicht vorstrukturierter KNN auf der Ebene der realisierten Rechenoperationen zu reflektieren. Zum einen liegt der wesentliche Schritt im Autoadaptionsprozess nicht in den formalen Rechenoperationen im Rahmen derer etwa Aktivierungsfunktionen ausgewertet werden, sondern in der resultierenden Adaption der Struktur des KNN. Zum anderen adaptieren künstliche neuronale Netze ihre Struktur wie beschrieben im

⁷⁵ Der finale Zustand des vorstrukturierten KNN ist auch deshalb nicht von besonderem Interesse, weil er so gut wie sicher nicht intelligibel ist.

⁷⁶ Die Autoadaption erfolgt nicht im Hinblick auf ein Optimalitätskriterium und kann nicht ohne Weiteres als eine Suche beschrieben werden. Die entsprechenden MLA suchen keine Ordnung in den Rohdaten, sondern reagieren lediglich systematisch auf Irritationen.

Rahmen der Aktualisierung einer großen Zahl abstrakter Verbindungsgewichte, die zwar einen schrittweise auswertbaren Algorithmus darstellen, aber keine für den Nutzer verständlichen Variablen symbolisieren. Eine nützliche Intuition besteht darin, sich vorzustellen, dass ein KNN weniger wie ein Algorithmus rechnet, als dass es wie ein elektrischer Schaltkreis schaltet. Insgesamt machen es die fehlende **Erwartbarkeit**, Steuerbarkeit und sogar Reflektierbarkeit des Einsatzes von nicht vorstrukturierten KNN beziehungsweise der entstehenden Strukturvorschläge zumindest problematisch, bei Netzen von Informationstechnik oder überhaupt von Technik zu sprechen.

3.2. Abgrenzung von etablierten Begriffsverwendungen

Das übergeordnete Ziel dieser Arbeit besteht in der techniknahen Vorbereitung weiterer Analysen maschinellen Lernens. Entsprechend soll zunächst geklärt werden, welche Schwierigkeiten der Diskussion im Wege stehen, um dabei festzuhalten, in welche Richtungen sie weiterentwickelt werden könnte oder sollte. Im Vorherigen wurde am Beispiel der nicht vorstrukturierten KNN bereits dargestellt, dass die Rede vom Zweck eines MLA nur für einen Teil der Artefakte angemessen ist. Diese Vorgehensweise soll nun auf eine Reihe von weiten Perspektiven und Begriffen ausgedehnt werden, bezüglich derer man Informationstechnologie und speziell maschinell lernende Algorithmen diskutieren kann. Die ausgewählten Begriffe sollen dabei ein Kontinuum von Perspektiven abdecken und werden so gesetzt beziehungsweise verstanden, dass sie disjunkte, aber vergleichbare Ansätze darstellen⁷⁷.

Das Ziel dieses Abschnittes und der späteren Betrachtung technikphilosophischer Entwürfe besteht darin, die Problematiken der Beschreibung maschinellen Lernens sichtbar zu machen. Die Absicht und der Anspruch dieses Abschnittes liegen dementsprechend nicht darin, erschöpfend zu zeigen, welche der begrifflichen Perspektiven abzulehnen sind beziehungsweise welche Anpassungen der jeweiligen Begriffsbildungen vorgenommen werden müssen, um MLA angemessen behandeln zu können. Derartige Anpassungen sind an dieser Stelle verfrüht, da noch nicht ausreichend geklärt ist, was überhaupt beschrieben werden soll. Zunächst muss, im Sinne des konstruktiven Erkundens von begrifflichen Sackgassen, geklärt werden, welche Aspekte maschinellen Lernens von einfacheren Beschreibungsmöglichkeiten jeweils vernachlässigt werden. Jeden Versuch einer frühen Begriffsbildung würden implizite Vorgaben der gewollten beziehungsweise zu vermeidenden Assoziationen begleiten. Gerade beim Einsatz von maschinellem Lernen können die äußerst vielfältigen Kontexte des Einsatzes der Artefakte sehr unterschiedliche Redeweisen erfordern oder zumindest nahelegen. Genau derartige Diskussionen maschinellen Lernens in spezifischen Kontexten oder unter speziellen Forschungsfragen sollen mit Hilfe dieser Arbeit vorbereitet und unterstützt werden. Vor diesem Hintergrund sollen zunächst die Schwierigkeiten in der Darstellung von MLA identifiziert werden. Begriffliche Perspektiven, die partiell angemessen zur Beschreibung von MLA erscheinen oder zumindest ein diesbezügliches Potenzial aufweisen, werden zusätzlich als mögliche Anknüpfungspunkte markiert.

3.2.1. Überraschung

Die erste Frage soll sein, ob der Autoadaptionsprozess oder die entstehenden Strukturvorschläge die Nutzer **überraschen** oder **enttäuschen** können. Die **Nichterwartbarkeit** der Ergebnisse nicht vorstrukturierter MLA bedingt sofort, dass zumindest für diesen Teil des maschinellen Lernens nicht sinnvoll von Überraschungen gesprochen werden kann. Die Nutzer werden von den entstehenden Strukturvorschlägen allenfalls verwirrt. Falls sich eine Überraschung einstellt, bedeutet das, dass die Nutzer im Vorfeld bereits Vorannahmen über die mögliche Interpretierbarkeit des Strukturvorschlages getroffen hatten und nun feststellen, dass diese Vorannahmen bezüglich des Nutzens falsch oder genau richtig sind. Im Normalfall werden solche Vorannahmen in der Praxis nicht getroffen, da auch für eine ungefähre Vorwegnahme der Kategorisierung in den meisten

⁷⁷ Aus diesem Grund werden die Begriffe als per Redekonvention gesetzt verstanden und nur ihr Nutzen in Hinblick auf maschinelles Lernen betrachtet. Lediglich der Zusammenhang der Begriffe untereinander wird thematisiert. Die einzelnen Begriffe können jeweils auch anders gefasst werden, aber an dieser Stelle ist in erster Linie der Perspektivenraum von Interesse, der von den Begriffen insgesamt aufgespannt wird.

Fällen einiger Aufwand notwendig ist. Es könnte argumentiert werden, dass eine **Erwartung** enttäuscht wird, wenn der entstandene Strukturvorschlag beziehungsweise die Kategorien sich nicht nutzbringend interpretieren lassen. Das wäre allerdings gleichbedeutend mit der Aussage, dass die Nichterfüllung eines Wunsches die Enttäuschung einer Erwartung bedeutet. Mangels Erwartung ist entsprechend der Begriff der Enttäuschung genau wie der Begriff der Überraschung im Kontext nicht vorstrukturierter KNN zunächst nicht angemessen.

Die Betrachtung der Überraschung ist dennoch in sich nützlich, da sie verdeutlicht, dass die Interpretation eines Strukturvorschlages eine Irritation des Nutzers voraussetzt. Erst mit dem Konstatieren des Vorliegens von etwas Neuartigem oder Erwähnenswertem und dem Versuch der Reproduktion oder Analyse dieser Beobachtung wird theoretisches Wissen benötigt oder gewonnen. Das bedeutet, dass der Strukturvorschlag den Nutzern nur eine Wahrnehmung oder Perspektive eröffnet und die Erkenntnis des Vorliegens eines Falles – also von „etwas“ – nachgelagert und keineswegs zwingend mit dem Autoadaptionsprozess nicht vorstrukturierter KNN verbunden ist. Eine interessante Einschränkung bezüglich der nachgelagerten Gewinnung von theoretischem Wissen liegt darin, dass diese Art der Struktursuche in sehr komplexen Kontexten – wie der Mathematik – auf Schwierigkeiten stößt. Dort kann eine nutzbringende Interpretation der Strukturvorschläge nur gelingen, wenn die resultierenden Kategorien sich zu einer Vielzahl von theoretischen Einschränkungen konform verhalten. Kategorien, die diese Rahmenbedingungen nicht berücksichtigen, sind dann pauschal uninteressant. Ein Beispiel aus der Industrie für solch eine komplexe Umgebung ist das Design von Mikrochips, bei dem durch maschinelles Lernen Schaltpläne erstellt werden, die zu Beginn größtenteils schlicht nicht funktionieren (Koza et al. 1996). In diesem Beispiel wird in der Praxis **evolutionäres Lernen** eingesetzt, das jedoch die gleiche Unerwartbarkeit aufweisen kann wie nicht vorstrukturierte KNN.

Eine Möglichkeit, wie die Rede von Überraschung bei MLA von Nutzen sein kann, besteht darin, die **Entropie** als ein Maß für Überraschung oder Informationen zu nutzen. Ein MLA, das Entropie reduziert, hat jedoch notwendigerweise auch eine Zielfunktion, die diese Reduktion misst und Zustände beziehungsweise Teilmengen von Trainingsdaten mit niedriger Entropie vorzieht. Hier verdeutlicht sich, dass die Bestimmung eines Informationsgehaltes ein systematisch anderes Teilgebiet maschinellen Lernens im Blick hat als die Suche nach im Vorfeld unbekannten Strukturen.

3.2.2. Repräsentation

Die zweite betrachtete Begriffsbildung ist die der **Repräsentation**. Manche Algorithmen des maschinellen Lernens verarbeiten Eingaben, indem sie diese Trainingsdaten oder zumindest Zusammenhänge innerhalb der Trainingsdaten durch ihre Struktur repräsentieren. Ein Beispiel sind **Entscheidungsbäume**, die Strukturvorschläge erstellen, die die übergebenen Trainingsdaten abbilden und die dazu dienen, die Zugehörigkeit von neuen Eingabedaten zu bestimmten Kategorien möglichst schnell und transparent prüfen zu können. Künstliche neuronale Netze repräsentieren hingegen nicht die übergebenen Sensordaten, sie entwickeln lediglich im Rahmen des Autoadaptionsprozesses die Eigenschaft, strukturiert auf Eingaben zu reagieren. Zwar hängt die Art der Reaktion von den übergebenen Trainingsdaten ab, aber es ließe sich allenfalls sagen, dass KNN die Trainingsdaten in einer wahrnehmbaren Form **präsentieren** – eine Re-präsentation hingegen entsteht erst mit der Interpretation des Strukturvorschlages durch den Nutzer. Künstliche neuronale Netze ohne Vorstrukturierung erstellen somit keine Repräsentationen und keine Modelle von Wirklichkeit. Auch eine

Reduktion des Anspruches an die Strukturvorschläge auf die Forderung einer Repräsentation von Modellen schlägt fehlt, denn es fehlt ein Maßstab, bezüglich dessen die Nützlichkeit des Strukturvorschlages eingeschätzt werden könnte. Das **Toy Modeling**⁷⁸ kann in diesem Zusammenhang als illustratives Gegenbeispiel zu nicht vorstrukturierten KNN angesehen werden, da dort zwar ebenfalls keine maximale Wirklichkeitsnähe, sehr wohl aber eine maximale Nützlichkeit angestrebt wird. Ein wesentlicher Spezialfall der Präsentation von Rohdaten durch MLA liegt vor, wenn die Rohdaten vor Beginn des Autoadaptionsprozesses eine unstrukturierte Menge ohne Form darstellen und daher für den Nutzer nicht wahrnehmbar sind. Insbesondere auf diesem Spezialfall basiert die später ausgearbeitete Idee, maschinelles Lernen als eine technische Unterstützung bei der Entdeckung von **Welt** zu betrachten.

Die Rede von Repräsentationen ist eng verknüpft mit der in der Informatik verbreiteten Interpretation von Eingabedaten als Instanzen und von Trainingsdaten als Trainingsinstanzen. Aus den vorangegangenen Überlegungen ergibt sich, dass die Verwendung der Begriffe der Instanz und Trainingsinstanz im Kontext des maschinellen Lernens zumindest problematisch ist. Die Sensordaten, die KNN erhalten, instanziierten etwa nicht unbedingt eine Struktur⁷⁹ und sind zu Beginn des Prozesses lediglich verfügbar. Leicht umformuliert besteht der Wunsch der Nutzer beim Einsatz von nicht vorstrukturierten KNN darin, eine Struktur als durch die Sensordaten instanziiert erkennen zu können, um darauf aufbauend mit den vormaligen Rohdaten arbeiten zu können. Die Sensor- beziehungsweise Trainingsdaten, anhand derer der Autoadaptionsprozess durchgeführt wird, als Trainingsinstanzen zu bezeichnen, ist daher irreführend. Der Begriff **Training** – als verändernde Entwicklung durch die Verarbeitung von Reizen – passt in diesem Kontext hingegen recht gut. Der Autoadaptionsprozess eines künstlichen neuronalen Netzes kann daher sinnvoll als Entwicklung von systematischen Reaktionen auf Trainingsdaten verstanden werden.

Inszenierungen, stabile Präsentationen und Repräsentationsoptionen

In einigen Bereichen der **Nanotechnik** findet sich eine ähnliche Situation wie im maschinellen Lernen. Dort liegt ebenfalls eine neue Form von Technik vor, bei der nicht die beiden Grundoperationen des **Regelns** und des **Steuerns** zum Zuge kommen.

But in general [controlled placement of every component atom] is infeasible: in the worst case it could need the global control and choreography of the behaviour of every individual nanite. A more feasible approach is to exploit mainly local cooperation between suitably-programmed neighbouring nanites, possibly mediated by their shared local environment (which also more closely mirrors the way biological organisms grow).

(Milner et Stepney 2003)

Stattdessen wird von **medialer Steuerung** gesprochen (Wiegerling 2012), womit gemeint ist, dass Umgebungen **inszeniert** werden – in der Absicht zu beobachten, was aufgrund der

⁷⁸ Die gezielte Nutzung stark vereinfachter und aller Wahrscheinlichkeit nach falscher Grundannahmen, die dennoch das relativ beste Prognosemodell in einem komplexen oder weitgehend unerforschten Kontext bilden (Gottschalk-Mazouz 2012).

⁷⁹ Sensordaten können sowohl ein zufälliges Rauschen wiedergeben als auch klar unterscheidbare Kategorien beschreiben.

jeweiligen Manipulation entsteht⁸⁰. Das Konzept der Inszenierung und das der Präsentation stellen verwandte Ansätze dar, die beide das Potenzial zu haben scheinen, in der Analyse nicht vorstrukturierter MLA beziehungsweise KNN von Nutzen sein zu können. Der Status des maschinellen Lernens als Technik wird im Weiteren in Hinblick auf diese beiden Konzepte weiter aufgearbeitet. Repräsentationen in ihrer klassischen Fassung wurden bereits als adäquate Beschreibung des Agierens nicht vorstrukturierter KNN und der resultierenden Strukturvorschläge abgelehnt. KNN dieser Art sind mangels Bezugsinstanz keine Mechanismen, die etwas repräsentieren, da eine Repräsentation eine zweistellige Relation darstellt – dennoch findet im Rahmen der Autoadaption des KNN an die Strukturen der unbekannten Rohdaten offenbar etwas statt. Wenn etwa ein KNN Rohdaten in Form eines Clusters präsentiert, können die Nutzer beginnen, über diesen Cluster nachzudenken und ihn zu interpretieren. In diesem Fall wird formal im Rahmen eines Präsentationsereignisses mittels eines Strukturvorschlages etwas dargestellt, das ohne das entsprechende MLA nicht präsentiert würde und zu dem sich die Nutzer in ein Verhältnis setzen können. In den meisten Fällen maschinellen Lernens basiert die Option der Nutzer, sich in ein Verhältnis zum Strukturvorschlag setzen zu können, auf einer Eigenschaft, die zufällige Präsentationen nicht haben – einer gewissen Form von **Stabilität** oder zumindest potenzieller Stabilität. Ein Beispiel der hier gemeinten Stabilität außerhalb des maschinellen Lernens besteht in der Höhlenmalerei (Hubig 2011). In Hubigs Beschreibung unterscheidet sich das an der Wand festgehaltene Tier in der Stabilität der Darstellung von einem zufälligen Blick auf das entsprechende Tier. Diese Stabilität ermöglicht es, sich mit der Angst vor dem Tier auseinander zu setzen oder Jagdstrategien zu planen⁸¹. Genau so können von einem MLA erzeugte Cluster als stabile Visualisierungen eines sonst flüchtigen Blickes auf chaotische Rohdaten beschrieben werden. Sobald eine Präsentation stabil ist, können die Betrachter sich zu ihr in ein Verhältnis setzen, sie interpretieren und somit die Präsentation zu einer Re-Präsentation werden lassen – durchaus im buchstäblichen Sinne von „Re“, da sich die Präsentationen in diesem Fall gleichsam rückwärts auf etwas bezieht. Angenommen, die Rohdaten erscheinen aus systemischen Gründen kontingent und es ist zunächst völlig offen, ob die Strukturvorschläge interpretiert werden können, dann liegt eine **höherstufige Erwartung** vor, wenn dennoch angenommen wird, dass Strukturvorschläge entstehen, zu denen sich die Nutzer in ein Verhältnis setzen können. Solange die prinzipielle Option der Nutzer, sich in ein Verhältnis zu den Strukturvorschlägen zu setzen, als gegeben angenommen wird, kann dem Autoadaptionsprozess entsprechend höherstufig eine Erwartbarkeit von **Repräsentationsoptionen** unterstellt werden. Insgesamt entsteht so ein in den meisten Kontexten des maschinellen Lernens anwendbares Gefälle von Repräsentationen über stabile Präsentationen hin zu Repräsentationsoptionen. Wenn dieses Gefälle zum Einsatz kommen kann, scheinen die basalen Kategorien zur Charakterisierung eines Umgangs mit Technik höherstufig verwendet auch die neuartigen Formen von Techniken noch angemessen modellieren zu können. Womit nicht in Frage gestellt wird, dass eine objektstufige Verwendung von Begriffen wie etwa Überraschung und Erwartbarkeit, wie sie die klassische Steuer- und Regelungstechnik betreffen, zurückzuweisen ist. Als ein objektstufiges Kriterium zur Sicherstellung der Anwendbarkeit des genannten Gefälles könnte angeführt werden, dass die Kenntnis einer ausführlichen Heuristik es ermöglicht, die Wahrscheinlichkeit angeben zu

⁸⁰ Details zu der Idee hinter diesem Begriff finden sich bei Hubig (Hubig 2006, S. 181ff, S. 185ff, S. 253).

⁸¹ Hier wird angenommen, dass zumindest bei einem Teil der Höhlenmalereien die Absicht (analog zum maschinellen Lernen) nicht primär darin liegt „Kunst“ zu erstellen.

können, mit der der Einsatz eines bestimmten MLA zum Erhalt von Repräsentationsoptionen führt. Eine höherstufige Erwartung darauf zu reduzieren, dass die Erstellung beliebiger Strukturvorschläge antizipiert wird, ist formal möglich. Beispielsweise wird bei einer nicht vorstrukturierten Clusteranalyse zwar nicht vorgegeben, *wovon* Cluster erstellt werden oder *wie* die Cluster aussehen sollen, aber dennoch ist sehr wahrscheinlich, *dass* Cluster erstellt werden. Eine solche formale Reduktion auf die Erwartung einer prinzipiellen Funktionalität des MLA ist jedoch wenig hilfreich, wenn die entstehenden Strukturvorschläge kontingent erscheinen und sich jeder Interpretation entziehen. Gegebenenfalls sind etwa die erstellten Cluster nicht reproduzierbar beziehungsweise stabil. In diesem Fall unterscheiden sie sich nicht von einer rein zufälligen Trennung der Rohdaten. Das bloße Entstehen der Cluster oder allgemein von Strukturvorschlägen legt somit nicht ohne Weiteres einen Zweck der zugrunde liegenden MLA frei – vor allem im Hinblick auf den Extremfall der spontanen, völlig unvorbereitet agierenden MLA, der im Folgenden im Fokus stehen wird.

Kontexte, in denen keine Stufe des Gefälles von Repräsentationen über stabile Präsentationen hin zu Repräsentationsoptionen anwendbar ist, erscheinen zunächst als Sonderfälle maschinellen Lernens, deren Relevanz fraglich ist. Diese Einschätzung ist einerseits zutreffend, da Forscher und Entwickler in der Informatik sich nicht primär daran orientieren, ob ihre Prototypen aus interdisziplinärer Perspektive kategorisch neuartige Technik erzeugen oder darstellen⁸². Andererseits ist speziell für die Technikphilosophie gerade das Auftreten neuer Technik von zentralem Interesse. Eine hohe Auflösungs- und Aussagekraft in diesem Bereich – oder zumindest eine möglichst präzise Beschreibung des Extremfalls der völlig spontanen, völlig unvorbereitet agierenden MLA – ist dementsprechend durchaus erstrebenswert. Die Idee einer höherstufigen Stabilität ist ein solcher Versuch, besitzt jedoch das notwendige Auflösungsvermögen noch nicht in ausreichendem Maße. Der Versuch einer präziseren Fassung des Begriffs der Höherstufigkeit, stellt dementsprechend einen möglichen Anknüpfungspunkt an die vorliegende Arbeit dar. Der Bedarf an einer Auflösung des unstrukturierten maschinellen Lernens entstammt darüber hinaus keinem rein theoretischen Interesse der Technikphilosophie. Der Grund ist, dass MLA, die die genannten Extremfälle realisieren, zwar nicht performant sind, aber in der Praxis durchaus eingesetzt werden. Ein praktisches Beispiel für den Einsatz von evolutionärem Lernen findet sich in der Optimierung eines Überschall-Kampfflugzeugs.

The vehicle was optimized using a genetic algorithm (GA) Queuing Multi-Objective Optimizer (QMOO) (Leyland, 2002). This type of algorithm is more suited to a hypersonic vehicle optimization problem than a gradient-based optimizer due to the nature of the problem: highly constrained, mixed-integer variables, and non-linear spaces in the solution space would cause a gradient based method to often get stuck in local optima.

(Smith 2009, S. 44)

Insgesamt gelingt mit Hilfe des Gefälles von Repräsentationen über Inszenierungen und stabile Präsentationen hin zu Repräsentationsoptionen die erste Beschreibung eines Spannungsbereichs beziehungsweise einer Übergangszone zwischen optimierenden,

⁸² Strukturvorschläge, die zur Erkennung von Verkehrszeichen eingesetzt werden können, oder ähnliche Anwendungsfälle lassen sich auf Basis von nicht vorstrukturierten KNN nur schwer realisieren.

repräsentierenden MLA wie den Entscheidungsbäumen und quasi-chaotischen MLA wie dem evolutionären Lernen.

3.2.3. Experiment

Die nächste mögliche Perspektive besteht darin, den Autoadaptionprozess eines MLA beziehungsweise den resultierenden Strukturvorschlag als eine Form von **Experiment** innerhalb der Informationstechnik zu betrachten. Ein Experiment ist hierbei nicht gedacht als Erschließung von Ähnlichkeiten, wie etwa zwischen realen Messungen und den Ergebnissen einer **Simulation**. Stattdessen ist die in Frage kommende Experimentform diejenige, bei der nicht mehr gesteuert und geregelt, sondern analog zur **medialen Steuerung** bewusst eine unübersichtliche Umgebung zugelassen oder gesucht wird. Am ehesten entspricht dies Rheinbergers Konzept von **epistemischen Dingen** übertragen in die Informatik⁸³.

Während technische Dinge eine angebbare Funktion in der Herstellung anderer Dinge haben oder als Dinge selbst zum Gebrauch und Verbrauch bestimmt sind, sind epistemische Dinge Erkenntnisgegenstände, also Objekte, an denen oder über die wir Wissen gewinnen wollen.

(Rheinberger et Herrgott 2001, S. 61)

Epistemische Dinge zeichnen sich bei Rheinberger dadurch aus, dass der Nutzer beziehungsweise der Forscher sie nicht gezielt beeinflusst und steuert, sondern chaotischen Einflüssen unterwirft und die Resultate beobachtet. Nicht vorstrukturierte KNN auf diese Weise als Erkenntnisgegenstände zu betrachten, an denen Wissen gewonnen werden kann, erscheint zunächst möglich. Eine solche Übertragung lässt sich vornehmen, indem das KNN und der Kontext des Autoadaptionprozesses mit dem Experimentalsystem in Analogie gesetzt werden und der Strukturvorschlag als Entsprechung des epistemischen Dings betrachtet wird. Das Problem an dieser Analogie ist, dass bei Rheinberger das epistemische Ding als dasjenige, was noch unbekannt ist und erforscht werden soll, durch das technische Ding darzustellen ist. Zwar sind dies bei Rheinberger keine ontologischen Begriffe und die Leitdifferenz findet sich in der Unterscheidung zwischen einem KNN, das einen Autoadaptionprozess durchlaufen hat und dem Strukturvorschlag, den das KNN darstellt, wieder⁸⁴, allerdings sind die Operationsmechanismen des KNN nicht zugänglich. Somit ist es schwierig, diesbezüglich die Analogie des technischen Dinges aufrecht zu erhalten, insbesondere, wenn auch der Kontext des KNN nicht zugänglich ist, wie etwa bei chaotischen Rohdaten⁸⁵. Die Nutzung der Perspektive Rheinbergers bietet sich dennoch insofern an, als er das Experimentalsystem metaphorisch als Maschine zur Herstellung von Zukunft bezeichnet und darauf abzielt darzustellen, dass Experimente so angelegt sind, dass ihr Ausgang offen ist. Hier findet sich eine sehr starke Parallele zu KNN, insbesondere weil die Offenheit bei KNN noch weiter gefasst ist als in den meisten klassischen Experimenten. Auch unstrukturierte KNN erhalten etwa Vorgaben bezüglich der Aktivierungsfunktionen und der Ausgangstopologie, somit

⁸³ Rheinberger betrachtet auch den produktiven Umgang mit Nichtwissen, der im Weiteren ebenfalls eine Rolle spielen wird.

⁸⁴ Diese Unterscheidung wird in der Diskussion der Nichttrivialität von MLA noch klarer ausgearbeitet.

⁸⁵ Es ist ebenfalls nicht sinnvoll, das KNN selbst als epistemisches Ding zu betrachten, außer der Nutzer ist daran interessiert, den internen Operationsmechanismus des KNN zu erschließen.

werden sie trotz der fehlenden Strukturierung sehr wohl so angelegt, dass ihr Ergebnis offen bleibt. Gerade die Möglichkeit, Artefakte dieser Art anlegen zu können, stellt eine zentrale Errungenschaft des maschinellen Lernens dar.

Unabhängig von der Analyse der Übertragbarkeit von Rheinbergers Begrifflichkeiten scheint der Begriff des epistemischen Dings durchaus eine zumindest partiell angemessene Annäherung an den Einsatz nicht vorstrukturierter KNN zu bieten. Auch unter Vernachlässigung der genannten Herausforderungen verbleibt dabei jedoch eine Schwäche, die vor einer Rede von einem Experiment im Zusammenhang MLA gelöst werden müsste. Die Schwäche dieser begrifflichen Perspektive liegt darin, dass mit Hilfe der MLA keine Effekte erzeugt werden sollen, die Wissensgewinn möglich machen, sondern dass Perspektiven gesucht werden, Rohdaten wahrzunehmen. Der fehlende Schritt aus Sicht der Nutzer ist primär die Beobachtbarkeit und nicht die daran anknüpfende Interpretation und der Wissensgewinn, die beide vom Nutzer allein realisiert werden. Im Beispiel der Analyse des Kundenstamms eines Supermarktes etwa wurden zwar bedingte Wahrscheinlichkeiten ermittelt, aber dass solche Wahrscheinlichkeiten prinzipiell mess- beziehungsweise formulierbar sind, ist keine neue Erkenntnis. Erst die Interpretation der Strukturvorschläge kann das Wissen über die Eigenheiten der Kunden erweitern. Demgegenüber inszeniert auch ein ungesteuert verstandenes Experiment, das keine Parameter testet, einen Prozess, um Effekte zu erzeugen deren Zustandekommen unklar ist, die aber allenfalls überraschend, aber nicht völlig unvorhergesehen auftreten. MLA können aufgrund ihrer hohen Geschwindigkeit und der aus der Virtualität resultierenden geringen Kosten der Prozessschritte eine extrem große Anzahl von autoadaptiven Iterationen durchlaufen. Im Fall der nicht vorstrukturierten künstlichen neuronalen Netze ist zunächst weder das Auftreten einer Überraschung noch überhaupt eines Effektes angestrebt beziehungsweise möglich. Das KNN als Struktur ist ebenfalls gerade nicht von Interesse. Der Nutzer richtet seine Aufmerksamkeit auf den Strukturvorschlag und dieser war vor dem Autoadaptionsprozess noch in keiner Weise greifbar. Sehr wohl möglich und für die weiteren Betrachtungen festzuhalten ist jedoch, dass Netze beim Design von Experimenten oder als ein „Schritt Null“ eines deutlich über den Autoadaptionsprozess hinausgehenden Experiments eingesetzt werden können. Ein KNN kann die Sichtbarmachung einer Umgebung übernehmen, woraufhin ein Experiment gegebenenfalls erst möglich wird. Mittels künstlicher neuronaler Netze kann in sehr vielen autoadaptiv erzeugten Strukturen gezielt auf ähnliche Weise experimentell improvisiert werden, um Informationen und daraus Wissen zu gewinnen. So können Netze durchaus genutzt werden, um die Produktivität im experimentellen Umgang mit Nichtwissen zu erhöhen, dennoch bilden sie dabei nur einen nachrangigen Teil des jeweiligen Versuchsaufbaus.

3.2.4. Zufallstechnik

Im Vorangegangenen wurden die systematische Repräsentation, die Repräsentationsoption und das ergebnisoffene Experiment in Bezug auf maschinelles Lernen betrachtet. Die Erwartbarkeit der Autoadaptionsprozesse und Strukturvorschläge kann entsprechend der bisherigen Systematik noch einmal reduziert werden, indem die Autoadaptionsprozesse als **Zufallstechnik** verstanden werden. Algorithmen des maschinellen Lernens sind jedoch noch immer Algorithmen und somit nicht in der Lage, einen völlig zufälligen Prozessschritt zu realisieren. Jeder echte Zufall kann prinzipiell nur aus unbekannten und kontingenten Sensordaten resultieren. Solch eine Kontingenz in den Sensordaten kann auch gezielt erzeugt werden, indem ein zufälliges Signal aus einer anderen Quelle mit Hilfe gewisser

mathematischer Methoden⁸⁶ über die Sensordaten gelegt wird. Auf diese Weise können die Messfehler der Sensordaten und damit deren Kontingenz künstlich beliebig erhöht werden. Diese Verfälschung der Sensordaten kann ohne Probleme so ausgeweitet werden, dass sich die vom MLA vorgeschlagenen Strukturen nur noch auf das überlagernde Störsignal beziehen und so keine Interpretation mehr zulassen.

MLA erstellen mitunter unerwartete Möglichkeiten, übergebene Trainingsdaten zu strukturieren, dieser Vorgang ist jedoch – bei Übergabe derselben Sensordaten – reproduzierbar. Der Autoadaptionsprozess, der den Strukturvorschlag entstehen lässt, ist gegebenenfalls nicht beobachtbar oder intelligibel, aber der Strukturvorschlag ist bei identischem Ausgangszustand und identischen Sensordaten gleichfalls identisch. Wie bereits angedeutet wurde, erschweren es Sensordaten mit großen zufälligen Anteilen den jeweiligen MLA sogar, Strukturen zu finden. Unabhängig von diesen Möglichkeiten ist es nicht das Ziel des maschinellen Lernens zu versuchen, Zufallszahlen oder Ähnliches zu erzeugen oder zu approximieren. Tatsächlich können KNN mittels übermäßig chaotischer Sensordaten gezielt zu Reaktionen gebracht werden, die den Symptomen menschlicher Schizophrenie ähneln (Hoffman et al. 2011)⁸⁷. Schizophrenie ist hier verstanden als eine irrationale aber nicht rein zufällige Reaktion auf Sensordaten⁸⁸. Die Idee hinter dieser Nutzung von KNN in der klinischen Forschung lag darin, die These zu prüfen, dass die Symptome von Schizophrenie beim Menschen durch eine Schwäche bei der Fähigkeit zur Filterung von irrelevanten Sinneseindrücken erzeugt werden. Unabhängig davon, ob diese These sich überhaupt mit Hilfe von KNN plausibilisieren lässt, ist der Gedankengang für das maschinelle Lernen von Bedeutung, weil auch hier häufig davon ausgegangen wird, dass eine ungebremste Messung von Allem ein Rauschen ergibt oder dass die gegebenenfalls verbliebenen Restsignifikanzen zumindest unzugänglich bleiben. Diese Annahme wiederum ist von Relevanz in der hier geführten Diskussion, da im maschinellen Lernen implizit auch die Gegenrichtung gedacht wird. Implizit scheint dort immer vom Vorliegen einer interpretierbaren Struktur ausgegangen zu werden, wenn die Rohdaten systematisch ein beschränktes Gebiet beschreiben. Möglicherweise liegt hierin auch ein Teil der Motivation, Trainingsdaten pauschal als Trainingsinstanzen zu bezeichnen. Diese Gleichsetzung und die verbundenen Gedankengänge werden hier nicht geteilt, daher ist es wichtig, die pauschale Annahme des Vorliegens einer interpretierbaren Struktur innerhalb der Informatik als einen möglichen Ursprung der Gleichsetzung festzuhalten.

3.2.5. Unfall

Insgesamt stellt das Konzept der Zufallstechnik, wie beschrieben, zu hohe Ansprüche an die Autoadaptionsprozesse maschinell lernender Artefakte. In einer letzten begrifflichen Abgrenzung soll entsprechend eine Verortung des nicht vorstrukturierten maschinellen Lernens *zwischen* dem Experiment und der Zufallstechnik versucht werden. Ein Begriff, der diesen Zwischenbereich betrachtet, ist derjenige des **Unfalls**. Das Auftreten eines

⁸⁶ Typischerweise eine Faltung etwa mit einem weißen Rauschen.

⁸⁷ Hier reicht es nicht, dem KNN verrauschte Daten zu übergeben, sondern es muss auch sichergestellt werden, dass das KNN das Rauschen der Trainingsdaten als echte Schwankungen der Messwerte versteht. Die Gegenmaßnahmen des KNN müssen gezielt umgangen werden, das heißt, es handelt sich um eine künstlich erzeugte Situation, die nicht typisch für das Verhalten von KNN oder MLA ist.

⁸⁸ Das heißt, Klassifizierungen werden zwar nach einer nicht nachvollziehbaren und häufig inkonsistenten Systematik vorgenommen, aber sie werden vorgenommen.

Strukturvorschläge unter völligem Fehlen einer **Erwartung** durch die Nutzer stellt ein Beispiel für einen Un-Fall im Sinne eines Nicht-Falles dar⁸⁹. Allerdings setzt diese Rede vom Un-Fall als Nicht-Fall die Existenz eines Falles voraus, von dem sich der Un-Fall als nicht erwartet abgrenzt. Die Charakteristik der Ausnahme oder der Abweichung trifft jedoch auf Strukturvorschläge im Kontext kontingent erscheinender Rohdaten gerade nicht zu – eben dieses Fehlen eines ursprünglichen Falles erfordert meist überhaupt erst den Einsatz des MLA. Entsprechend ist die Beschreibung eines Strukturvorschlages als Unfall ebenfalls nicht sinnvoll möglich. Allerdings kann das Konzept des Unfalls im Rahmen eines deutlich über den direkten Einsatz von MLA hinausgehenden Experimentes sehr wohl Verwendung finden. Experimente der in Abschnitt 3.2.3 genannten Art können gezielt darauf ausgelegt sein zu prüfen, ob neu auftretende Rohdaten Instanzen einer bereits bekannten Struktur oder strukturloses Rauschen darstellen, beziehungsweise ob auf Basis der Rohdaten eine neue Struktur denkbar wird, die systematisch aber unerwartet von einem der bereits bekannten Parameter abweicht. Der letztere Fall kann als ein Unfall verstanden werden, der es wert wäre untersucht zu werden, um das mögliche Entstehen einer nützlichen oder gefährlichen Struktur prognostizieren zu können. Aufbauend auf der Darstellung des Zusammenspiels von maschinellern Lernen und Experimenten können MLA im Rahmen der Prognose von Unfällen Daten gezielt unvoreingenommen beziehungsweise erwartungslos aufnehmen. Mittels wiederholt ähnlich improvisierter Experimente kann dann versucht werden, den komplexen sich ständig ändernden Raum der Unfälle zumindest ein wenig im Blick zu behalten.

⁸⁹ Die Bezeichnung eines Un-Falles als Nicht-Fall entspricht nicht dem Ursprung des Wortes. Stattdessen bezeichnete ein Unfall zunächst lediglich einen Fall oder einen Fall mit negativen Konsequenzen: „unfall entspricht dem einfachen fall [...] dessen übles ergebnis betonend [...]“ (Grimm 1971).

3.3. Ein Schritt zurück: MLA als Informationstechnik und Technik

Die vorangegangene Abgrenzung von etablierten Begrifflichkeiten als Mittel zur Beschreibung des maschinellen Lernens hat die Problematik der Beschreibung der Autoadaptionprozesse deutlich werden lassen. Gleichzeitig deutet der prinzipielle Charakter der Herausforderungen bei der Beschreibung von Teilbereichen des maschinellen Lernens an, dass der Ursprung der Schwierigkeiten noch genauer bestimmt werden muss. Zunächst sollen in diesem Abschnitt die grundlegenden Annahmen zur Verortung maschinellen Lernens in Frage gestellt werden. Anschließend wird im nächsten Abschnitt die Intuition, mit der von maschinellem Lernen die Rede ist, problematisiert und es werden Problembegriffe gesucht, die helfen sollen, die Schwierigkeiten zu lokalisieren.

Eine der zu hinterfragenden grundlegenden Annahmen ist, dass es sich auch bei nicht optimierend-repräsentierenden Formen maschinellen Lernens, wie nicht vorstrukturierten **KNN**, um **Informationstechnik** im wörtlichen Sinn handelt. Mit der Übergabe von kontingenten Rohdaten werden keine Informationen an das KNN übermittelt, die Eingabe dieser Form von maschinellem Lernen ist demnach unverdächtig, Informationsverarbeitung zu betreiben. Im Rahmen des Autoadaptionprozesses steht die Adaption einer dem MLA eigenen Struktur im Vordergrund. Die Details dieser Frage werden in der Diskussion der Nichttrivialität noch weiter beleuchtet, zu diesem Zeitpunkt ist jedoch schon klar, dass der Fokus der Autoadaption in der Manipulation von Strukturen wie KNN liegt und nicht in der Trennung von Rohdaten. Gleichwohl wird solch eine Trennung in vielen Fällen von den entstehenden Strukturvorschlägen geleistet, da diese auf eine systematische Art und Weise auf Eingaben reagieren und dabei die Eingaben in Klassen aufteilen. Die Klassen stellen jedoch Unterschiede dar, die zunächst keinen Unterschied machen.

Information ist ein Unterschied, der einen Unterschied macht. [Ökologie des Geistes]

(Wikipedia Contributors 2013, Gregory Bateson)

Eine reine **Separierung** von Daten ist durchaus bei vielen Formen von Technik zu finden, da Technik fast immer systematisch auf Eingaben reagiert und somit formal Eingaben klassifiziert⁹⁰. Die von nicht vorstrukturierten KNN vorgenommenen Klassenbildungen erstellen Unterschiede, die erst einen Unterschied machen, wenn sie im Nachgang durch den Nutzer interpretiert wurden. Nicht vorstrukturierte KNN dienen entsprechend nicht der Verarbeitung von Informationen, sondern können allenfalls formal, im Rahmen der Separation von Klassen, als Ermöglicher von Informationen angesehen werden. Insgesamt stellt die Klassifizierung von Rohdaten keine Form von Informationsverarbeitung dar und die Rede von Informationstechnik ist zumindest problematisch.

Nachdem exemplarisch am Beispiel der nicht vorstrukturierten KNN gezeigt wurde, dass MLA sich nicht ohne Weiteres als Informationstechnik beschreiben lassen, ist eine weitere Grundannahme zu hinterfragen: ist ein MLA überhaupt Technik und wenn es keine Technik ist, was genau ist es dann? Was ist die Seite am zugrunde liegenden **Algorithmus**, die dazu führt, dass, wenn man einen lernenden Algorithmus lange mit sich und einigen Sensordaten alleine lässt, etwas eher Atechnisches entsteht? Im Bisherigen wurden die

⁹⁰ Im trivialsten Fall werden die Eingaben nach zulässigen und unzulässigen Eingaben getrennt.

Autoadaptionsprozesse und die zugrunde liegenden Ideen und Konzepte beschrieben und für eine technikphilosophische Analyse vorbereitet, es blieb jedoch offen, inwiefern es sich überhaupt um Technik handelt – speziell wenn vorausgesetzt wird, dass im Technikbegriff eine Art von Nutzen- oder Zweckperspektive impliziert ist. Die Betrachtung ist jedoch nicht auf diesen Technikbegriff beschränkt, etwa treten auch in Hinsicht auf Technikbegriffe, die auf Wiederholbarkeit und Ursächlichkeit basieren, Probleme auf, wenn Strukturvorschläge lediglich unvorhergesehene Korrelationen präsentieren. Aus der Diskussion kann die Intuition gewonnen werden, dass die **Nichterwartbarkeit** der Strukturvorschläge von nicht vorstrukturierten KNN und ähnlichen MLA bedingt, dass die *einmalige* Durchführung eines Autoadaptionsprozesses und die Erstellung eines *isolierten* Strukturvorschlag tendenziell einen atechnischen Vorgang darstellt⁹¹, der gegebenenfalls als algorithmisches Gestalten oder eine Art von maschineller **Kunst** beschrieben werden kann⁹².

Die Motivation MLA zu entwickeln, liegt jedoch – wie häufig beim Einsatz von Computern – in der äußerst hohen Arbeitsgeschwindigkeit der Artefakte. MLA sind in der Lage, den Nutzern in kurzer Zeit extrem viele mögliche Struktur- beziehungsweise Wahrnehmungsvorschläge zu ansonsten kontingenten, nicht beobachtbaren Rohdaten anzubieten. Sofern die Sensordaten schnell genug verfügbar sind, können MLA weiterhin sehr effizient mit externen Analysemethoden gekoppelt werden, da der gesamte Autoadaptionsprozess elektronisch abläuft. Im Kontext dieser über maschinelles Lernen deutlich hinausgehenden Anwendungen *scheint* in der Praxis der Wunsch der Nutzbarkeit von MLA, wie bereits diskutiert, über heuristische Erfolgswahrscheinlichkeiten höherstufig wieder zu einer **Erwartung** zu werden. Zum einen ist jedoch die Rolle des maschinellen Lernens in einem solchen Versuchsaufbau noch nicht geklärt und zum anderen wird im Rahmen solch einer Konstruktion das Ziel wieder zu einem Wunsch, wenn die Erschließung absolut unbekannter Daten eingesetzt wird. Ein Beispiel wäre der Versuch, in den auf der Erde messbaren Sensordaten Strukturen zu finden, die sich als außerirdische Kommunikationsversuche verstehen lassen. Hierzu gibt es keine erfolgreichen Versuche aus der Vergangenheit und es gibt keinerlei Erwartung, wie solch eine Kommunikation aussehen könnte⁹³. Unabhängig von der Frage nach den Änderungen durch den massenhaften Einsatz von Autoadaptionsprozessen wurde in der bisherigen Diskussion am Beispiel der KNN gezeigt, dass Teile des maschinellen Lernens sich als Typ von Technik insofern deutlich von der klassischen Technik unterscheiden, als keine Steuer- und Regelungsprozesse identifiziert oder gar unterschieden werden können.

Die Frage ist, was all das auf den Technikbegriff zurückwirkend bedeutet. Maschinelles Lernen scheint in Teilen einen Phänomenbereich darzustellen, der dazu zwingt, Begriffe neu zu entwerfen. Die Betrachtungen der etablierten Begriffe stellten ein Pendeln zwischen Abgrenzung und Präzisierungsgewinn dar. Einerseits konnte etwa eine optionale Zwecklosigkeit festgestellt werden und andererseits werden MLA in der Praxis häufig gerade deshalb eingesetzt, weil sie objektstufig aufgrund einer Heuristik gerade doch einen Nutzen versprechen. Dieses Pendeln sollte insbesondere aufzeigen, dass auch elaborierte Vorschläge dafür, was unter Technik verstanden werden kann, jeweils auf mindestens einer der Ebenen

⁹¹ Der Zusammenhang zwischen Erwartung und Technik wird von Kaminski aufgearbeitet (Kaminski 2010).

⁹² Dieser Intuition wird im Weiteren nicht systematisch gefolgt, allerdings wird sie punktuell wieder thematisiert werden.

⁹³ Darüber hinaus werden sich in der notwendigerweise endlichen Menge der in der Praxis erhobenen Rohdaten zwangsweise Strukturen finden lassen. In der Praxis aufgefundene Strukturen stellen dementsprechend sehr wahrscheinlich keine Kommunikationsversuche, sondern Zufallseffekte dar.

ausgehebelt werden – auch wenn sie der Betrachtung dabei durchaus eine gewisse Präzisierung ermöglichen. Wenn es bei diesen Fragen nur um die Determination von Algorithmen ginge, wäre die Problemlage mit logischen Mitteln vergleichsweise einfach auflösbar – das Interessante ist jedoch, dass dem nicht so ist und dass die entsprechenden Artefakte sich einer Beschreibung entziehen. Zusammengefasst stellt sich die Frage, welche Faktoren dazu führen, dass manche MLA agieren wie beispielsweise nicht vorstrukturierte KNN. Erst die Formulierung einer These zur Beantwortung dieser Frage erlaubt eine zielgerichtete, konstruktive Diskussion der Selbstorganisation im maschinellen Lernen. Im nächsten Abschnitt werden mit der Neugier und der Vor-Struktur zwei Problembegriffe freigelegt, die eine konstruktive Denkrichtung für solch eine Diskussion aufzeigen.

3.4. Suche nach Problembegriffen

Im Bisherigen wurde herausgearbeitet, dass sich ein Teilbereich des maschinellen Lernens einer Beschreibung konsequent entzieht. Der entsprechende Bereich wurde etwas unscharf umschrieben als die Menge derjenigen MLA, die wie nicht vorstrukturierte KNN agieren. Neben den KNN wurden Teile des evolutionären Lernens als Beispiele genannt, während repräsentierende Entscheidungsbäume als Gegenbeispiel dienten. Der Bereich der MLA, die nicht vorstrukturierten KNN ähneln, wurde darüber charakterisiert, dass die entsprechenden MLA auf Basis besonders geringer Vorgaben systematisch auf Reize beziehungsweise Irritationen reagieren. Dieser Abschnitt soll ausgehend vom Hintergrundwissen aus dem ersten Hauptteil Problembegriffe freilegen, die eine differenzierte Beschreibung und Diskussion der einzelnen Teilbereiche des maschinellen Lernens ermöglichen. Die Hauptziele bestehen darin, die Frage wieder aufzugreifen, was an einer Diskussion maschinellen Lernens technikphilosophisch besonders interessant erscheint und zu motivieren, warum eine Unterscheidung von MLA und anderen IT-Artefakten in vielen Kontexten lohnenswert sein kann. Ein wesentliches Nebenziel wird darin bestehen, mit Hilfe der Gewinnung von Fragestellungen und Problembegriffen eine Prüfung aktueller technikphilosophischer Entwürfe in Hinblick auf deren Eignung für die Diskussion maschinellen Lernens vorzubereiten.

Die Diskussion wird in zwei Schritten erfolgen. In einem ersten Schritt werden die Problematik und der Diskussionsbedarf dargestellt. Es soll dabei ein Unwohlsein beziehungsweise eine Irritation bezüglich gewisser Rede- und Denkweisen im Kontext maschinellen Lernens aufkommen. Die im Fokus stehende Rede ist dabei die Verortung von MLA als technische Artefakte⁹⁴ im Gegensatz zu einer Beschreibung als quasi-intentionale, subjektartige Selbstorganisatoren. Die Irritation besteht darin, dass Anmutungen von Subjektivität bei MLA aufkommen. Zunächst ist nicht klar, wie mit dieser überraschenden Redeweise umgegangen werden kann. Im zweiten Schritt wird der Irritation mit Hilfe der **Daseinsanalytik** Heideggers und dessen Weltbegriffes gefolgt, und mögliche Ursachen dieses Unwohlseins werden aufgespürt und als Problembegriffe isoliert. Weder werden hierbei MLA als Subjekte dargestellt, noch werden sie vollständig im Artefaktbegriff aufgehen.

Zusammengefasst stellt sich die Frage, in welcher Hinsicht *genau* maschinell lernende Artefakte subjektartig zu sein scheinen. In welchen Hinsichten sind MLA nur Mittel, die die Nutzer bei ihrem Beziehen auf die Welt nutzen können, und inwiefern scheinen sie etwas zu sein, das selbsttätig eine eigene Welt entdeckt.

3.4.1. Quasi-Intentionalität als Ausgangspunkt der Suche

Die Perspektive der Informatik auf MLA wurde in der Einleitung und im ersten Hauptteil bereits diskutiert. Demgegenüber besteht in Bezug auf maschinell lernende Artefakte jedoch auch eine alltägliche, quasi-intentionale Redeweise⁹⁵. Es besteht die starke Intuition, dass MLA einerseits als Artefakte technisch hergestellt sind und andererseits einen quasi-intentionalen

⁹⁴ Die Frage, ob MLA als technische Artefakte, Systeme oder als Maschine modelliert werden, spielt hier nur eine nachgeordnete Rolle. Zunächst soll die Diskussion maschinellen Lernens möglich gemacht werden, die Diskussion technikphilosophischer Perspektiven erfolgt im nächsten Abschnitt.

⁹⁵ Auch in Kontexten des **UbiComp** oder **smarter** Artefakte taucht diese Redeweise auf, selbst wenn maschinelles Lernen nicht oder nur sehr nachrangig zum Einsatz kommt. Diese Fälle könnten gesondert betrachtet werden, die Redeweise basiert dort jedoch meist auf der Undurchsichtigkeit und den fehlenden Schnittstellen beziehungsweise Spuren der Artefakte und nicht auf der im Folgenden freigelegten Intuition.

Aspekt aufweisen. Im vorherigen Abschnitt wurde bereits dargestellt, dass die Rede von Informationstechnik MLA einen Umgang mit Information und damit eine subjektähnliche Qualität zuschreibt. Alltagssprachliche Beispiele für die quasi-intentionale Qualität informatischer Technik in der Lebenswelt finden sich exemplarisch in den folgenden Formulierungen:

- Das Recommender-System empfiehlt mit Vorliebe Produkt A.
- Das Artefakt erstellt Kategorien für die Trainingsdaten und wählt anschließend für ein Testdatum B eine besonders passende Kategorie aus.
- Das Artefakt stört sich am Eingabedatum C, weil es nicht den bisherigen Daten entspricht. Das Artefakt passt sich an die veränderte Umwelt an, um diese Irritation zu verringern.

Derartige Formulierungen können, insbesondere im Zusammenhang mit der generell recht metaphorischen Beschreibung von Software und Algorithmen, die Frage verbergen, wer oder was in welcher Weise aktiv ist, wenn ein KNN sich an Eingabedaten „stört“. Die Formulierungen erzeugen ein gewisses Unwohlsein, da sie auf der Intuition basieren, dass MLA den Anschein einer Quasi-Intentionalität erwecken – ohne dass diese Intuition begründet oder gar diskutiert werden kann. Die diesen quasi-intentionalen Formulierungen zu Grunde liegende Intuition soll im Weiteren beschrieben und mit Hilfe von Problembegriffen genauer gefasst werden. Es wird der Versuch unternommen, Problembegriffe zu finden, die möglichst klar festhalten, in welcher Hinsicht MLA sich nicht ohne Weiteres als technische Artefakte verorten lassen. Zwar soll geprüft werden, inwiefern die quasi-intentionale Redeweise zulässig ist, allerdings soll dabei nicht gefragt werden, inwieweit MLA Subjekte *sind*, sondern nur wieso sie so *erscheinen*. Gleichzeitig werden sich MLA nicht vollständig auf klassische Artefakte reduzieren lassen, ohne das intuitive Phänomen der Eigentätigkeit beziehungsweise der Quasi-Intentionalität aufzugeben beziehungsweise zu verlieren. Das Ziel bei der Suche nach Problembegriffen wird sein, festzuhalten, wo genau das **Changieren** zwischen Artefakt und Subjekt entsteht. Was sind die Momente, die dieses Changieren provozieren beziehungsweise motivieren?

Welt als Grundintuition

Die Grundintuition hinter der nachfolgenden Suche nach hilfreichen Beschreibungen des Changierens ist, dass der **Weltbegriff** einen sinnvollen ersten Versuch darstellt, einen Problembegriff zu finden, der eine konstruktive Diskussion von MLA erlaubt. Die Rede von den Reizen, die ein nicht vorstrukturiertes KNN aus seiner Umgebung aufnimmt und von einer reaktiven Anpassung an die sich verändernde „Umwelt“ scheint einen interessanten Kern zu haben. Die Beschreibung der von nicht vorstrukturierten KNN erstellten Strukturvorschläge als maschinelle Weltbezüge und des Autoadaptionsprozesses als eine Form des Erkennens von Welt löst gerade die bereits beschriebene Irritation aus. Gleichzeitig scheinen diese Redeweisen genau dasjenige zu beschreiben, was *manche* MLA von anderen IT-Artefakten unterscheidet. Welt als Problembegriff beschränkt die Diskussion somit einerseits erfolgreich auf das maschinelle Lernen und erzeugt andererseits weiterhin das genannte Changieren. Der Begriff der Welt liefert somit, in seiner Thematisierung als Problembegriff, einen Einstieg in die Suche nach einer präziseren Modellierung des Changierens. Eine Theorie, mit deren Hilfe Problembegriffe für eine detailliertere Modellierung gesucht werden sollen, muss neben dem Weltbegriff möglichst detaillierte Unterscheidungsmöglichkeiten verschiedener Aspekte der Subjektivität analysieren. Auf diese Weise wird die Möglichkeit

geschaffen, einen speziellen Aspekt der Subjektivität zu finden, der das Phänomen beinhaltet, das zur beschriebenen Irritation geführt hat und der gleichzeitig weniger umfassend ist als der Weltbegriff.

Eine Ausarbeitung solch einer differenzierten Betrachtung der Subjektivität mit Berücksichtigung der Welt stellt die Daseinsanalytik Heideggers in *Sein und Zeit* (Heidegger 1927) dar. Lesern, denen *Sein und Zeit* fremd ist, reicht eine kurze Einführung in die Denkweise Heideggers, um der Argumentation folgen zu können. Kompakte Darstellungen der im Weiteren benötigten Begriffe, die nur geringe beziehungsweise keine Lektürekenntnisse erfordern, bieten Herrmann oder Weber (Heidegger et Herrmann 1989; Weber 2010, S. 12ff). Eine ausführlichere Aufarbeitung findet sich in Wikibooks (Wikibooks Contributors 2012, *Sein und Zeit*). Leser aus der Informatik können *Weisheit - Wissen - Information* (Gloy et zur Lippe 2005) als Tertiärliteratur für die Arbeit mit Wikibooks nutzen. Heideggers Perspektive bietet sich an, weil die Beschreibung des Daseins der Versuch ist, gezielt *kein* übergeordnetes Subjekt zu postulieren, das handelt oder einen Willen aufweist, sondern gewissermaßen eine Subjektphilosophie ohne Subjekt zu betreiben. Heidegger versucht gerade die alltägliche Redeweise im Umgang mit Dingen und das zielgerichtete Verhalten des Menschen mit einem neuartigen Zugriff zu erfassen. Zwar ist die philosophiegeschichtliche Neu- beziehungsweise Andersartigkeit seines Ansatzes im Kontext maschinellen Lernens ohne Bedeutung, allerdings beschädigt das nicht das Auflösungsvermögen, das Heideggers Begrifflichkeiten bei der Betrachtung des Daseins ermöglichen. Seine detaillierte Analyse verschiedener Aspekte beziehungsweise Momente des Daseins bietet eine Vielzahl möglicher Problembegriffe. Einige dieser Optionen werden im Weiteren im Kontext des maschinellen Lernens betrachtet und es wird bestimmt, welche Begriffe das beobachtete Phänomen des Changierens noch erzeugen und gleichzeitig einen möglichst spezifischen Bereich der Subjektivität beschreiben. An dieser Stelle ist explizit nicht Heideggers weitergehende Analyse der Bedingungen der Möglichkeit der von ihm festgestellten Weltbezüge von Interesse – genauso wenig wie sein Verständnis von Technik. Das heißt, es geht nicht um eine Heidegger-Interpretation, da keine stabile Analogie zwischen dem Dasein und MLA erstellt oder konstatiert werden soll. Die Nutzung der Analyse Heideggers dient dazu, genauer zu bestimmen, welche Momente des Daseins ein MLA zu besitzen *scheint*. Die Diskussion ist damit eröffnet und nicht abgeschlossen. Die Betrachtung von Heideggers Daseinsanalytik soll dementsprechend kein Problem lösen, da noch keine Problemstellung identifiziert oder gar formuliert wurde. Insofern soll bei der Suche nach Problembegriffen soweit möglich vermieden werden, implizit Lösungen zu investieren. Der Einsatz von *Sein und Zeit* (Heidegger 1927) erleichtert eine solche ergebnisoffene Suche, da Heideggers Fokus von der Betrachtung quasi-intentionaler Strukturen bei Artefakten weitgehend unabhängig ist. Weiter entspricht die Idee einer ergebnisoffenen Betrachtung des Phänomens des Unwohlseins methodisch einem phänomenologischen Zugriff, wie etwa demjenigen Heideggers.

Der zweckrationale Charakter der folgenden Betrachtung von *Sein und Zeit* (Heidegger 1927) soll zunächst kurz an einem der bereits genannten Beispiele detailliert werden.

- Das Artefakt stört sich am Eingabedatum C, weil es nicht den bisherigen Daten entspricht. Das Artefakt passt sich an die veränderte Umwelt an, um diese Irritation zu verringern.

Ausführlicher beschrieben, besagt diese Aussage, dass einige MLA Eingabedaten prüfen und feststellen, ob sie dem widersprechen, was das MLA „erwartet“. Diese „Erwartung“ kann einer vom Nutzer vorgegebenen Zielfunktion entstammen oder sich aus der jeweils aktuellen Struktur des auf Basis der bisherigen Eingabedaten erstellten Strukturvorschlags ergeben. Erst wenn ein Eingabedatum auftritt, das dieser Erwartung widerspricht, „lernt“ das MLA, indem der Autoadaptionsprozess eingeleitet wird. In den Worten Heideggers ausgedrückt scheint es, als ob für einige MLA von einer **Zuhandenheit** der Eingabedaten gesprochen werden kann. In quasi-intentionaler Redeweise formuliert, scheinen die Eingabedaten im Autoadaptionsprozess einiger MLA nur dann eine Rolle zu spielen, wenn sie **auffälliges, aufsässiges** oder **aufdringliches Zeug** darstellen und nicht einfach zuhanden sind. Diese Redeweise dient hier nur dazu, die Irritation zu vergegenwärtigen und das Changieren zwischen der Verortung eines MLA als Artefakt und der Zuschreibung einer Form von Subjektivität vorzuführen. Speziell für den Teilbereich der später als zielorientiert bezeichneten MLA kann jedoch durchaus konstruktiv davon gesprochen werden, dass der Anschein einer Zuhandenheit der Eingabedaten entsteht. Im Gegensatz zu den später als neugierig beschriebenen MLA weisen zielorientierte Artefakte klare Vorstrukturierungen durch den Nutzer auf und *scheinen* entsprechend ein Konzept von der Verwendbarkeit von Eingabedaten zur Erreichung ihrer Ziele zu haben.

Die Beschreibung von Eingabedaten als Zuhandenes zeigt einen der Punkte, an denen das Changieren in der Rede von maschinellem Lernen deutlich wird. Die Suche nach der Motivation einer Rede von einer Zuhandenheit von Eingabedaten führt sofort zu einer Vielzahl von anknüpfenden Fragen, wie der nach der **Bewandtnisganzheit**, in der ein MLA vom Nutzer eingesetzt wird.

- Können die Eingabedaten für den Nutzer des MLA vielleicht tatsächlich als ein Zuhandenes betrachtet werden, solange das MLA nicht auf die Daten reagiert?
- Modifizieren MLA die Bewandtnisganzheit und wenn ja, tun sie das transparent?
- Sind MLA Teil des **In-der-Welt-seins** der Nutzer oder der Programmierer der Artefakte und wenn ja, auf welche Weise?

Diese Fragen sollen hier nicht diskutiert werden, denn sie besitzen für die weitere Analyse des maschinellen Lernens keine große Relevanz. Die genannten Fragen spielen eher im Kontext des UbiComp eine Rolle als bei einer Diskussion des maschinellen Lernens und dessen Selbstorganisationsprinzipien. Allerdings ist es bereits ein Mehrwert, zwischen diesen beiden Diskursen klarer trennen zu können. An dieser Stelle soll lediglich veranschaulicht werden, wie die Freilegung von Problembegriffen eine Diskussion initiieren kann, allerdings tritt dabei auch die größte Herausforderung bei der Nutzung der Daseinsanalytik zum Vorschein. Die Hauptschwierigkeit liegt darin, dass die **Momente** des Daseins in Heideggers Modellierung sehr stark voneinander abhängen⁹⁶. Die einzelnen Momente lassen sich dementsprechend nicht zufriedenstellend isoliert voneinander zur Beschreibung eines anderen Kontextes heranziehen. Das heißt, Heideggers Perspektive kann allenfalls genutzt werden, Problembegriffe zu identifizieren, aber nicht dazu, diese im Kontext des maschinellen Lernens zu diskutieren. Demgegenüber wird ein Gewinn aus der Betrachtung des Heidegger'schen Begriffsapparates darin bestehen, gezielt Problembegriffe zu identifizieren, die nützliche

⁹⁶ Das Selbst kann überhaupt nur als Selbst angesprochen werden, wenn es die Angst vor dem Tod hat. Die Angst vor dem Tod vereinzelt, erst dadurch kann man sich selber zum Gegenstand machen und so fort.

Assoziationen aufweisen. Die starken expliziten Abhängigkeiten zwischen den Momenten des Daseins haben für die Diskussion den Vorteil, dass sie die Problembegriffe in gewissem Maße gegen implizite unzulässige oder verunklarende Assoziationen immunisieren. Assoziationen können im Kontext der präzisen Begriffe Heideggers sehr schnell auf ihre Zulässigkeit geprüft werden, wodurch ihre Grenzen deutlich werden. Zusammengefasst bietet Heidegger vielfältige und präzise Begriffe, um das Unwohlsein über die Rede von Quasi-Intentionalität im Kontext von MLA, mit Hilfe von Problembegriffen positiv und negativ einzugrenzen und zu **lokalisieren**.

Eigentätigkeit als alternative Grundintuition

Ein Nebenziel der Rede von Welt wird darin bestehen, das Zustandekommen der Strukturvorschläge daraufhin zu untersuchen, inwiefern es einen selbsttätigen Charakter hat und ob die MLA durch vorgegebene Zwecke, durch eigene und damit nichtmenschliche Zwecke oder durch keinerlei Zwecke bestimmt sind⁹⁷. Eine alternative Grundintuition, die diesem Nebenziel eine deutlich höhere Priorität zuordnet, wurde von Richter und Kaminski entwickelt (Richter et Kaminski 2013) und wird im Folgenden kurz dargestellt. Die Alternative besteht darin, nicht den Weltbegriff, sondern die **Eigentätigkeit** als begrifflichen Anker der Suche nach Problembegriffen zu nutzen. Im vorangegangenen Abschnitt zu Informationstechnik wurde gefragt, was an einem lernenden Algorithmus dazu führt, dass etwas eher Atechnisches entsteht, wenn man den Algorithmus mit sich und einigen Sensordaten allein lässt. Die Form dieser Frage führt eher in Richtung einer Eigentätigkeit als einer Welt des MLA. Scheinbar sind MLA, ähnlich wie Subjekte, in der Lage, eigentätig auf Anforderungen zu reagieren und ihre Voraussetzungen quasi-autonom zu verändern⁹⁸. Das ist zunächst überraschend und es ist unklar, wie intentional diese Vorgehensweise ist. Die Beobachtung einer durch ein Ding vermittelten Strukturierung von Welt kann jedoch auch unabhängig von maschinellern Lernen gemacht werden.

I call it the law of the instrument, and it may be formulated as follows: Give a small boy a hammer, and he will find that everything he encounters needs pounding

(Kaplan 1964, S. 28)

Für einen Jungen mit einem Hammer in der Hand zerfällt die Welt in die Klassen der hämmerbaren und der nicht-hämmerbaren Dinge – wobei Kaplan betont, dass die erste Klasse für ein Kind deutlich größer ist⁹⁹.

Die Betrachtung der Eigentätigkeit als Grundintuition würde dementsprechend zu einer deutlich anderen Diskussion maschinellen Lernens führen. Nicht zuletzt würde sich vor solch einem Hintergrund Heideggers Daseinsanalytik weniger anbieten, um die Diskussion zu eröffnen. Mittel treten für ihn nur durch die drei Modi der Auffälligkeit, Aufdringlichkeit und

⁹⁷ Auf diese Frage wird vor allem bei der Diskussion des Worumwillen in Abschnitt 3.4.2 eingegangen.

⁹⁸ Diese Beobachtung wird in der nachfolgenden Suche nach Problembegriffen wieder aufgegriffen, wenn davon die Rede ist, dass MLA ihre Vorstruktur adaptieren.

⁹⁹ Diese Form von Strukturierung ist nur übertragbar auf die Betrachtung von Strukturvorschlägen und nicht generell auf die Analyse von MLA. Der Hammer hat einen Hintergrund in einer Bewandtnisganzheit, die einem MLA nicht ohne Weiteres zugeschrieben werden kann. Dieser Punkt wird in der Suche nach Problembegriffen noch ausgearbeitet.

Aufsässigkeit in das In-der-Welt-sein eines Daseins und zuvor macht die Frage, ob Mittel oder Zuhandenes eine Eigentätigkeit aufweisen oder nicht, keinen Sinn. Die Frage wäre demnach, inwiefern Eigentätigkeit für das Dasein bei Heidegger eine Rolle spielt. Der Bezug auf eine Eigentätigkeit der Subjektivität würde eine neue Heidegger-Interpretation erfordern, die analysiert, inwiefern Eigentätigkeit möglicherweise einen Kern dessen berührt, was Heidegger als Dasein versteht. Eine Idee hierzu könnte darin bestehen, Eigentätigkeit als Kontrapunkt zu einem passiv ausgelieferten Dasein zu denken und Entwerfen als vorausblickende Tätigkeit und als tatsächlich handelnden Umgang mit der Welt zu betrachten. Wenn das Selbst-Sein ein Verhalten ist, das sich zu sich selbst verhält, ist damit in dieser Interpretation echte Aktivität gemeint – gedankliche, handelnde, planende Auseinandersetzung mit den Momenten, innerhalb derer sich das Dasein jeweils befindet¹⁰⁰.

Neben der Nutzung der Daseinsanalytik wäre auch eine Eröffnung der Diskussion mit Hilfe anderer philosophischer Perspektiven denkbar, soweit sie die genannten Voraussetzungen erfüllen. Ein Beispiel für eine in Hinblick auf die Idee der Eigentätigkeit recht naheliegende Alternative wäre der Einsatz einer **Medienphilosophie**. Eine Medienphilosophie würde sich insbesondere anbieten, weil die von den MLA erstellten Strukturvorschläge in späteren Abschnitten durchaus als weltvermittelnd modelliert werden. In dieser Vermittlung lassen sich die Strukturvorschläge je nach Zielstellung durchaus zufriedenstellend als Medium modellieren – allerdings wird dies für die zugrunde liegenden MLA nur sehr eingeschränkt gelten. Weiter würde die Nutzung einer solchen Perspektive implizit vermuten oder gar voraussetzen, dass MLA als Medien verstanden werden können und wäre damit weniger ergebnisoffen als eine Nutzung der Daseinsanalytik. Es wäre in diesem Fall im Nachgang schwierig zu erkennen, inwiefern spezielle MLA tatsächlich gut modelliert werden können und an welchen Stellen die Problembegriffe noch nicht detailliert genug sind, um eine konstruktive Auseinandersetzung mit dem maschinellen Lernen zu gewährleisten.

Der Fokus des weiteren Abschnittes liegt auf der Arbeit mit dem Weltbegriff, es birgt jedoch einen Mehrwert kurz festzuhalten, auf welche Schwierigkeiten eine Betrachtung von MLA als Medien zur Vermittlung von Welt stoßen würde: Die eingangs beschriebene Irritation speiste sich unter anderem aus Formulierungen wie der Verortung von Strukturvorschlägen als Weltbezüge und der Beschreibung von deren Erstellungsvorgang als eine Form des Erkennens der Welt. Weiter wurde beobachtet, dass die für den Nutzer undurchsichtige, autoadaptive Verarbeitung von Rohdaten durch MLA Strukturvorschläge entstehen lässt, die auf noch zu klärende Art beim Nutzer Erschlossenheit von Welt zu **induzieren** scheinen. Eine naheliegende Frage ist, ob hier nicht Welt *vermittelt* wird. MLA stellen kein maschinelles Dasein dar und weisen entsprechend weder selbst eine Welt auf, noch wird ihnen Welt vermittelt. Das heißt, es scheint zunächst nur eine dritte Position zu verbleiben, nämlich diejenige eines Dings, zu dem sich das Dasein verhält. Tatsächlich füllen die von MLA erstellten Strukturvorschläge diese dritte Position auch aus. Die MLA allerdings *erstellen* diese dritte Position erst, die gleichwohl ohne ein interpretierendes Dasein gerade keine „dritte Position“ ist. Ein zentraler Aspekt bei der Beschreibung maschinellen Lernens ist das nutzerseitige **Sich-in-ein-Verhältnis-Setzen** zu Strukturvorschlägen. Diese vom Nutzer initiierte Relation zu Strukturvorschlägen lenkt seinen Blick auf einen neuen, aber immer

¹⁰⁰ Das Hauptproblem wäre, dass in dieser Interpretation gegebenenfalls ein Subjektbegriff investiert werden müsste. In Heideggers Begrifflichkeiten kann jedoch gerade nicht ohne detaillierte Begründung davon gesprochen werden, dass es „etwas“ gibt, das sich zu den vorgegebenen Möglichkeiten verhält und dann auswählt und abwägt.

schon durch die Interpretation gegebenen Aspekt von Welt. Der Begriff der **Induktion** bezieht sich gerade darauf, dass der neue Aspekt von Welt nicht von einem Subjekt mittels eines Mediums entdeckt wird, sondern dass der Zugriff auf Welt, im Moment der nutzerseitigen Wahrnehmung des Strukturvorschlags, durch den Strukturvorschlag immer schon gegeben ist¹⁰¹. Die Interpretation des Strukturvorschlags durch den Nutzer ist immer schon erfolgt. Die Rede von Induktion soll hierbei verdeutlichen, dass der relevante Prozess vollständig auf der Seite des Menschen abläuft und lediglich die notwendigen Ressourcen vom MLA zur Verfügung gestellt werden – dass sich der Nutzer jedoch auch nicht dagegen verwehren kann, den Strukturvorschlag zu interpretieren, wenn er sich dazu entschließt, ihn zu betrachten. Der Strukturvorschlag ist in gewisser Weise immer schon interpretiert, verursacht diese Interpretation jedoch nicht. Die Schwierigkeiten der Beschreibung einer Funktion von Strukturvorschlägen als einer Form von induktivem Medium sollen hier nicht weiter analysiert, sondern als möglicher Anknüpfungspunkt festgehalten werden. In dieser Arbeit werden die den Strukturvorschlägen zugrunde liegenden MLA im Fokus stehen und sollen entsprechend gerade nicht auf ihre Erzeugnisse reduziert werden. In Abschnitt 3.6 wird untersucht werden, inwiefern die von MLA erstellten Strukturvorschläge den Nutzer bei der Erstellung von Welt *unterstützen* können. Diese Frage lässt sich jedoch in diesem Abschnitt im Rahmen der Daseinsanalytik nicht sinnvoll formulieren, ohne Strukturvorschläge – durchaus begründet – als reines Zeug zu verorten und der Diskussion somit ihre Relevanz zu nehmen.

Methodik der Suche

Das dargestellte Unwohlsein führt zum Verlangen nach einer Diskussion, die erklärt, in welchen Aspekten genau MLA dem Dasein oder dem klassischen bewusstseinsphilosophischen Subjekt zu ähneln scheinen. Die folgenden Betrachtungen werden zur Ermöglichung dieser Diskussion prüfen, welche der begrifflichen Modellierungen aus der Daseinsanalytik im Kontext des maschinellen Lernens das eingangs erwähnte Changieren auslösen. Das heißt, es werden Begriffe gesucht, bezüglich derer MLA Anmutungen von Subjektivität aufzuweisen scheinen. Ambivalenzen sind dementsprechend willkommen und gesucht werden Begriffe, die nicht genau passen, deren Einsatz aber auch nicht unzulässig erscheint. Kurz gesagt, wird eine konstruktive Erhöhung des Unwohlseins angestrebt.

Ausgehend vom In-der-Welt-sein werden begriffliche Konzepte in ihre Momente beziehungsweise Teilaspekte aufgelöst und es wird geprüft, welche der Teilaspekte im Kontext des maschinellen Lernens ebenfalls eine konstruktive Irritation erzeugen. Diese Teilaspekte rücken anschließend nacheinander in den Fokus. Andere Teilaspekte, die in Hinblick auf MLA eindeutig bewertet werden können, werden nicht weiter betrachtet, da sie nichts zur Eröffnung einer Diskussion des maschinellen Lernens beitragen. Diese Vorgehensweise lässt eine „Kette“ von Begriffen entstehen, die immer genauer den Ursprung der Irritation lokalisieren. Diese Kette setzt sich aus Begriffen zusammen, die einerseits in *Sein und Zeit* Aspekte des In-der-Welt-seins beschreiben und andererseits im Kontext des maschinellen Lernens als Problembegriffe konstruktiv die Diskussion eröffnen können. Die entstehende Kette von Begriffen ist in der nachfolgenden Grafik kurz zusammengefasst. Weitere Konzepte Heideggers, wie dasjenige der Befindlichkeit und das der Rede, sind bewusst *nicht* in der Grafik enthalten, da sich herausstellen wird, dass sie sich nicht sinnvoll für die Diskussion maschinellen Lernens nutzen lassen.

¹⁰¹ Der Begriff der Induktion ist hier *nicht* in Abgrenzung von Deduktion, sondern in seiner technischen Bedeutung verwendet.

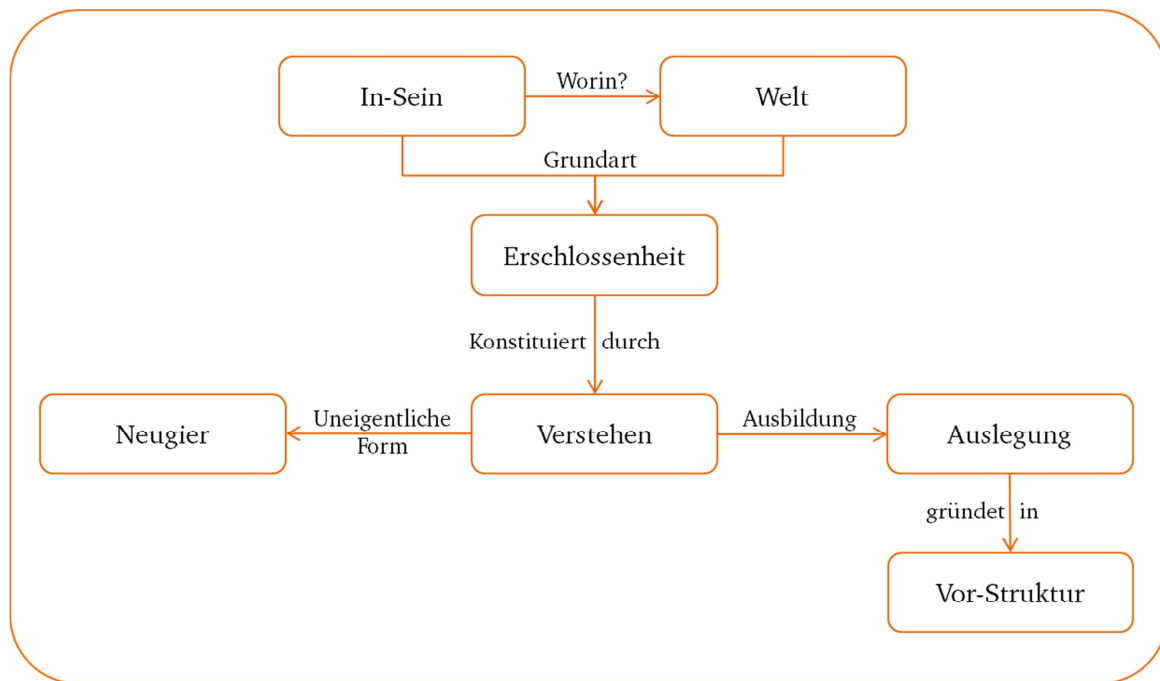


Abbildung 39: Übersicht der potenziellen Problembegriffe

Die beiden Problembegriffe, deren Verwendung durch die Beschäftigung mit Heideggers Perspektive motiviert werden soll, sind die **Neugier** und die **Vor-Struktur**. Beide Begriffe werden nichts final klären, sondern sich als geeignet erweisen, eine Diskussionsrichtung vorzugeben. Es soll und wird keine Parallelstruktur zu Heideggers Daseinsanalytik aufgebaut, sondern nur das beobachtete Phänomen begrifflich lokalisiert. Mit Hilfe der Neugier soll eine präzisere Abgrenzung des auf einer Form von Selbstorganisation basierenden Teilbereiches des maschinellen Lernens gegenüber dem auf mathematischer Optimierung aufbauenden Teilbereich ermöglicht werden. Die Vor-Struktur hingegen soll, aufbauend auf den Betrachtungen des ersten Hauptteils, eine Unterscheidung verschiedener Formen von Autoadaptivität auf der Ebene einzelner Algorithmen ermöglichen. Die Unterscheidung von MLA aufgrund unterschiedlicher Vor-Strukturen zielt darauf ab festzuhalten, dass verschiedene Formen von **Parametrisierungen** einer Lernstrategie nicht einfach unterschiedlich optimale Ergebnisse liefern. Stattdessen legen Parametrisierungen mitunter auch fest, wie „neuartig“ die Strukturvorschläge sein können, die das MLA aus einer Interpretation der entstehenden Strukturvorschläge gewinnen kann.

3.4.2. Suche nach Ursprüngen der Irritation

Die Grundintuition, dass der Weltbegriff ein sinnvoller Anker für die Suche nach dem Ursprung der genannten Irritation ist, kommt darin zum Ausdruck zu fragen, ob und wie MLA ihre eigene Welt erstellen, entdecken oder aufweisen können. Sollen MLA in Bezug auf das Welterkennen betrachtet werden, ist zunächst festzuhalten, dass Heidegger selbst das Erkennen von Welt explizit dem In-der-Welt-sein zuordnet.

Wenn wir jetzt danach fragen, was sich an dem phänomenalen Befund des Erkennens selbst zeigt, dann ist festzuhalten, daß das Erkennen selbst vorgängig gründet in einem Schon-sein-bei-der-Welt, als welches das Sein von Dasein wesentlich konstituiert.

(Heidegger 1927, S. 61)

Diesen Gedanken modifizierend soll betrachtet werden, inwieweit MLA als eine ausgezeichnete Art des Seienden beschrieben werden können, das eine mit dem In-der-Welt-sein verwandte Struktur aufweist.

In-Sein und Welt

Der erste Schritt besteht somit darin, die Frage zu diskutieren, welche Aspekte beziehungsweise Strukturmomente des In-der-Welt-seins prinzipiell nützlich sein können, um MLA eine dem Welterkennen verwandte Struktureigenschaft zuzuschreiben, und welche ohne Ambivalenz nur dem Dasein als einem speziellen In-der-Welt-sein zufallen. Hierzu macht Heidegger eine recht präzise Vorgabe.

Das In-Sein ist nach dem Gesagten keine »Eigenschaft«, die das Dasein zuweilen hat, zuweilen auch nicht, ohne die es sein könnte so gut wie mit ihr. Der Mensch »ist« nicht und hat überdies noch ein Seinsverhältnis zur »Welt«, die er sich gelegentlich zulegt. Dasein ist nie »zunächst« ein gleichsam in-sein-freies Seiendes, das zuweilen die Laune hat, eine »Beziehung« zur Welt aufzunehmen.

(Heidegger 1927, S. 57)

Ohne an dieser Stelle weiter darauf eingehen zu wollen, inwieweit es nur „ein“ Seinsverhältnis zur Welt geben kann, kann das In-Sein bei der Betrachtung von Welterkennen aus Heideggers Perspektive nicht sinnvoll vernachlässigt werden. Die Verwendung beider Begriffe erzeugt in Bezug auf MLA das konstruktive Unwohlsein und das Changieren zwischen Artefakt und Subjekt. Entsprechend werden im Weiteren das In-Sein und die Welt als die beiden für die Suche nach Problembegriffen relevanten Strukturmomente der Form des In-der-Welt-seins betrachtet.

Verstehen und Worumwillen

Im nächsten Schritt ist die Erschlossenheit des In-Seins und der Welt zu betrachten.

Erschlossenheit aber ist die Grundart des Daseins, gemäß der es sein Da ist. Erschlossenheit wird durch Befindlichkeit, Verstehen und Rede konstituiert und betrifft gleichursprünglich die Welt, das In-Sein und das Selbst.

(Heidegger 1927, S. 220)

Auf der Ebene der Erschlossenheit ist der Begriff der Befindlichkeit zu spezifisch auf das Dasein hin entwickelt, um im Zusammenhang mit MLA von Nutzen zu sein. Es findet sich

zwar eine Analogie für MLA, die Anzeichen des Changierens aufweist¹⁰², aber die Betrachtung dieser Analogie ist weder notwendig noch hilfreich, um MLA präzise diskutieren zu können. Für den Begriff der Rede gilt ähnliches: MLA – insbesondere diejenigen Artefakte, die nicht vorstrukturierten KNN ähneln – erwecken nicht den Anschein, ein Konzept von Bedeutung aufzuweisen¹⁰³. Nachdem Befindlichkeit bereits als zu spezifisch abgelehnt wurde, kann dementsprechend die Rede als „bedeutungsmäßige Gliederung der befindlichen Verständlichkeit“ (Heidegger 1927, S. 162) im Zusammenhang mit MLA ebenfalls als nicht relevant eingestuft werden. MLA weisen weder bezüglich der Befindlichkeit noch in Hinsicht auf die Rede ein Changieren auf, beide Begriffe können pauschal abgelehnt werden. Diese Nichtberücksichtigung von Befindlichkeit und Rede resultiert in der Fragestellung, ob der ursprünglichen Intuition mit der isolierten Betrachtung des Verstehens Genüge getan werden kann. Alternativ ist es möglich, dass mit Hilfe von Heideggers phänomenologischer Modellierung des Welterkennens prinzipiell nur zu den Strukturvorschlägen als beobachtbaren Ergebnissen von Autoadaptionsprozessen und präsentierenden „Vermittlern“ von Welt vorgedrungen werden kann.

Diese Befürchtung erweist sich als unbegründet, da die Rede von der Erschlossenheit des Verstehens sich sehr gut zur Erzeugung einer konstruktiven Irritation im Kontext des maschinellen Lernens einsetzen lässt. Die konkrete Vorgabe einer zu optimierenden Zielfunktion oder die Festlegung eines einzusetzenden Selbstorganisationsprinzips scheinen sich in Heideggers Worten als das **Worumwillen** des MLA beschreiben zu lassen. Dieses muss als ein gestiftetes Worumwillen betrachtet werden, da das MLA aufgrund seiner Autoadaptivität die vom Nutzer erhaltenen Vorgaben und Festlegungen gegebenenfalls verändert. Tatsächlich beschreibt die Möglichkeit einer Modifikation der eigenen Vorgehensweise auf Basis von Sensordaten genau den definierenden Aspekt aller MLA. Entsprechend ist es von besonderem Interesse, dass es den Anschein hat, als ob Heidegger ein vergleichbares Verhalten des Daseins mit dem Begriff des Verstehens beschreibt. Die quasi-intentionale Rede von MLA unter Bezug auf das Worumwillen soll aus diesem Grund im Folgenden etwas detaillierter dargestellt werden.

Im Worumwillen ist das existierende In-der-Welt-sein als solches erschlossen, welche Erschlossenheit Verstehen genannt wurde. Im Verstehen des Worumwillen ist die darin gründende Bedeutsamkeit miterschlossen. Die Erschlossenheit des Verstehens betrifft als die von Worumwillen und Bedeutsamkeit gleichursprünglich das volle In-der-Welt-sein. Bedeutsamkeit ist das, woraufhin Welt als solche erschlossen ist.

(Heidegger 1927, S. 143)

Ein MLA ist zwar initial vom Menschen parametrisiert und designt, in Heideggers Worten scheint ein MLA damit vom Dasein als Worumwillen geworfen zu sein, allerdings wird die ursprüngliche Motivation – etwa eine initialisierte Zielfunktion oder die Vorgabe einer bestimmten Form der Mutation – mitunter nicht final gesetzt, sondern nur **gestiftet**. Die Stiftung eines Worumwillen beschreibt, dass das MLA initialisiert, sich dann aber als

¹⁰² Diese Analogie findet sich in den faktischen konstruktiven Rahmenbedingungen jedes tatsächlich Seienden MLA, im abduktiven Bias der Lernstrategie und in den Initiationsparametern.

¹⁰³ Bedeutung und damit Sinn entsteht erst durch die Interpretation der Strukturvorschläge durch den Menschen.

Selbstzweck überlassen wird¹⁰⁴. Mit dieser Vorgabe eines Worumwillen weisen MLA eine Vergleichbarkeit zum uneigentlichen Sein des Daseins auf, wo das Worumwillen ebenfalls vorgegeben ist – beispielsweise durch eine Rolle. Dasein ist durch seine eigene Geschichte immer auch in seinem Sein und seinen weiteren Möglichkeiten bestimmt und ist insofern auch eingeschränkt im weiteren Entwurf. Der Mensch initialisiert das MLA und beobachtet dann ohne spezielle beziehungsweise spezifische **Erwartungen** die Autoadaptionsprozesse oder je nach Lernstrategie auch lediglich die Strukturvorschläge als deren Ergebnis¹⁰⁵. Dies gilt, wie bereits aus Perspektive der Informatik beschrieben, nicht für alle MLA, es finden sich auch Problemstellungen, bei denen die Möglichkeiten des MLA im Vorfeld bereits thematisch erfasst sind. Dies gilt etwa für die Stützvektormethoden. Anwendungen dieser Art können in Konsequenz den mathematischen **Optimierungsproblemen**¹⁰⁶ zugeordnet werden, und diese Fälle erwecken *nicht* den Eindruck, dass eine Form von Welterkennen vorliegt. In solchen Kontexten maschinellen Lernens gibt es, vereinfacht gesagt, nicht nur eine klare Vorgabe, wie die Lösung genau aussieht, sondern auch gleich einen konkreten Lösungsalgorithmus. Es tritt allerdings eine Vielzahl von Problemstellungen auf, bei denen die Lösung nicht oder nicht vollständig vorgegeben und bei denen das mit der Lösung verfolgte Ziel oder das über den Strukturvorschlag realisierte Selbstorganisationsprinzip ebenfalls nicht oder nur teilweise bekannt ist. Manchmal ist beispielsweise nur die Codierung bekannt, mittels derer Strukturvorschläge formuliert werden können, oder es ist nicht einmal das gegeben und stattdessen liegt nur eine riesige, chaotisch erscheinende Datenquelle vor, gegebenenfalls in Kombination mit einer vagen Vorstellung, was ein Ziel auszeichnen könnte. Entsprechend ist es wichtig, einerseits die Vorgaben des Nutzers, wie etwa die Wahl, welche Selbstorganisationsprinzipien eingesetzt werden, und andererseits dessen Motivation, überhaupt ein MLA einzusetzen, zu unterscheiden und die mögliche Unbestimmtheit beider Aspekte festzuhalten.

Insgesamt stellt das Verstehen einen nützlichen Zwischenschritt auf der Suche nach möglichen Problembegriffen dar. Allerdings muss der Begriff dennoch in seine Facetten aufgelöst werden, um den Detailgrad der betrachteten Begriffe noch einmal zu erhöhen und Aspekte abzuspalten, die die Diskussion mehr verunklaren als ermöglichen. Im Weiteren werden entsprechend drei Facetten des Verstehens aus Sicht des maschinellen Lernens genauer betrachtet: der Entwurf, die Neugier und die Auslegung. Als Resultat dieser Betrachtungen wird sich die Neugier als nützlicher Problembegriff erweisen, während die Auslegung in Hinsicht auf die Vor-Struktur weiter aufgelöst wird. Der Entwurf hingegen wird als Problembegriff zurückgewiesen werden.

Abweisung des Entwurfs

Die Autoadaptivität von MLA und die damit verbundene Ausrichtung von MLA auf die Möglichkeiten, auf Sensordaten zu reagieren, lassen zunächst die Vermutung entstehen, dass der Entwurf einen nützlichen Problembegriff darstellen könnte.

¹⁰⁴ Die Fähigkeit zur Autoadaption wird nicht nur gezielt zugelassen, sondern sogar gefördert.

¹⁰⁵ Das ist insbesondere interessant in Hinblick auf Technik als Erwartung (Kaminski 2010).

¹⁰⁶ Formal handelt es sich wie beschrieben dennoch um maschinelles Lernen, auch wenn die entsprechenden Probleme häufig bereits als „Optimierungsprobleme“ formuliert sind.

Warum dringt das Verstehen nach allen wesenhaften Dimensionen des in ihm Erschließbaren immer in die Möglichkeiten? Weil das Verstehen an ihm selbst die existenziale Struktur hat, die wir den Entwurf nennen. Es entwirft das Sein des Daseins auf sein Worumwillen ebenso ursprünglich wie auf die Bedeutsamkeit als die Weltlichkeit seiner jeweiligen Welt. Der Entwurfcharakter des Verstehens konstituiert das In-der-Welt-sein hinsichtlich der Erschlossenheit seines Da als Da eines Seinkönnens.

(Heidegger 1927, S. 145)

Im Kontext des maschinellen Lernens scheint es so, als ob das Autoadaptionsvermögen eines MLA auf seine Zielfunktion oder die einzusetzenden Selbstorganisationsprozesse hin ebenso wie auf bereits verarbeitete Roh- beziehungsweise Trainingsdaten ausgelegt wird. Die konstruktive Irritation, die durch die Rede eines Verstehens bei MLA entsteht, scheint bestehen zu bleiben. Allerdings stellt sich die Frage, ob die Rede von Möglichkeiten in Heideggers Sinn im Kontext von MLA sinnvoll oder problematisch ist. MLA verhalten sich nicht zu sich selbst, zeichnen sich also nicht durch das Strukturmoment des Selbst aus und beim Dasein steht im Gegensatz zu MLA immer der Freiheitsgedanke im Hintergrund¹⁰⁷. Heidegger selbst schreibt zu den Möglichkeiten und dem Entwurf das Folgende.

Der Entwurfcharakter des Verstehens besagt ferner, daß dieses das, woraufhin es entwirft, die Möglichkeiten, selbst nicht thematisch erfaßt. Solches Erfassen benimmt dem Entworfenen gerade seinen Möglichkeitscharakter, zieht es herab zu einem gegebenen, gemeinten Bestand, während der Entwurf im Werfen die Möglichkeit als Möglichkeit sich vorwirft und als solche sein läßt. Das Verstehen ist, als Entwerfen, die Seinsart des Daseins, in der es seine Möglichkeiten ist.

(Heidegger 1927, S. 145)

Die fehlende Freiheit von MLA und der fehlende inhaltliche Bezug von MLA zu den Rohdaten stellen dementsprechend ein Problem dar und bedingen, dass ein MLA nicht in vergleichbarer Weise in Bezug zu Möglichkeiten steht oder gar seine Möglichkeiten *ist*, wie das Dasein. In Hinblick auf die Rede von Möglichkeiten entsteht nicht der Anschein eines Changierens, sondern die Unterschiede sind klar. Die Schwierigkeiten, die auch schon das Verstehen aufwies, verstärken sich beim Entwurf so weit, dass der Einsatz des Entwurfes als Problembezug und die Rede von Möglichkeiten analogisierende Redeweisen wären, die mehr verunklaren würden, als sie an Problemen analytisch aufzeigen können¹⁰⁸. Insgesamt kann die Thematisierung des Entwurfes als Aspekt des Verstehens im Kontext des maschinellen Lernens als nicht hilfreich zurückgewiesen werden.

¹⁰⁷ Auch wenn man MLA gegebenenfalls nicht prognostizieren kann, kann man im Kontext des maschinellen Lernens nicht in einer sinnvollen Weise von Freiheit sprechen.

¹⁰⁸ Allenfalls zeigen diese Abweichungen eine Grenze auf, innerhalb derer Irritationen durch eine quasi-intentionale Redeweise als konstruktiv angesehen werden können.

Neugier als Problembegriff

Das Ziel der nachfolgenden Betrachtung von Neugier liegt darin darzustellen, dass einige MLA als in analytisch noch zu klärender Weise **neugierige Artefakte** bezeichnet werden können. Die Rede von Neugier soll hier, wie bereits beschrieben, nicht als Auflösung, sondern als Lokalisation eines Diskussionsbedarfs dienen. Anders ausgedrückt soll es darum gehen, dass es einen Mehrwert bietet, zu diskutieren, inwiefern nicht nur *man* neugierig sein kann, sondern auch *es*.

Was ist es um diese Tendenz zum Nur-Vernehmen? Welche existenziale Verfassung des Daseins wird am Phänomen der Neugier verständlich?

[...]

Das umsichtige Entdecken der Werkwelt hat den Seinscharakter des Ent-fernens. Die freigewordene Umsicht hat nichts mehr zuhanden, dessen Näherung zu besorgen ist. Als wesenhaft ent-fernende verschafft sie sich neue Möglichkeiten des Ent-fernens; das besagt, sie tendiert aus dem nächst Zuhandenen weg in die ferne und fremde Welt. Die Sorge wird zum Besorgen der Möglichkeiten, ausruhend verweilend die »Welt« nur in ihrem Aussehen zu sehen. Das Dasein sucht das Ferne, lediglich um es sich in seinem Aussehen nahe zu bringen.

(Heidegger 1927, S. 172)

Das Tendieren in eine ferne und fremde Welt, um sie nur in ihrem Aussehen zu sehen, stellt eine Formulierungsweise dar, die außerordentlich gut die Grundintuition wiedergibt, dass die Autoadaption von MLA auf Basis von Reizen und den Rückmeldungen zu vorangegangenen Autoadaptionen als eine Form von Reaktion auf Welt oder Suche nach Welt problematisiert werden kann. So scheinen zumindest diejenigen MLA, die *keine* expliziten Ziele verfolgen, durchaus die Fähigkeit aufzuweisen, in die ferne und fremde Welt zu tendieren. Im Folgenden wird die Begriffsverwendung noch ausgearbeitet, aber an dieser Stelle soll bereits festgehalten werden, dass mit der Neugier von MLA eine Quelle des eingangs erwähnten Unwohlseins identifiziert und als Problembegriff isoliert werden kann. Tatsächlich scheint dieser Begriff sogar eine definierende Eigenschaft nicht vorstrukturierter MLA zu beschreiben. Derartige MLA werden dementsprechend im Weiteren als **neugierige MLA** bezeichnet. Maschinell lernende Artefakte, die klare Vorgaben erhalten, welche Ziele zu erreichen und welche Parameter zu optimieren sind, werden hingegen im Folgenden als **zielorientierte Artefakte** bezeichnet¹⁰⁹.

Die Unterscheidung zwischen neugierigen und zielorientierten MLA soll es weiterführenden Diskussionen erlauben, sich auf die jeweils relevante Teilmenge von Artefakten zu beziehen. Die Trennung der beiden Teilbereiche des maschinellen Lernens ist jedoch nicht immer eindeutig. Zielorientierte MLA können etwa nicht pauschal so gedacht werden, als seien sie per nutzerseitiger Setzung mit einem definiten und invarianten Ziel ausgestattet. Neugier und Zielorientiertheit bilden stattdessen in ihrer Reinform die Enden eines Kontinuums. MLA zur

¹⁰⁹ Die Details zur Motivation einer Rede von Zielorientierung werden an späterer Stelle noch detailliert ausgeführt, insbesondere in der Diskussion der transklassischen Technik in Abschnitt 3.5.2.

Lösung mathematischer Optimierungsprobleme können als Beispiel für eine vollständig zielorientierte Ausprägung dienen, während ihre Struktur verändernde Kunstwerke (Brown et al. 2007) das andere Ende des Kontinuums und damit eine vollständig neugierige Ausprägung¹¹⁰ darstellen. Zusammenfassend werden viele neugierige und einige der zielorientierten MLA so konzeptioniert, dass sie sich mit Eingabedaten unbekannter oder fehlender Struktur beschäftigen können und dass im Rahmen des Autoadaptionprozesses ein Strukturvorschlag entsteht, der den Anschein erweckt, ein Umgang mit einer fernen und fremden Welt zu sein.

Sowohl für zielorientierte als auch für neugierige MLA gilt, dass sie im Rahmen eines Autoadaptionprozesses Strukturvorschläge erstellen, die für den Nutzer Rohdaten präsentieren und betrachtbar machen. Allerdings können speziell die Strukturvorschläge neugieriger MLA als Hinweise auf neue, von den Nutzern **un-vorher-gesehene**, Strukturen in den Eingabedaten betrachtet werden. Die von den MLA vorgeschlagenen Strukturen sind dabei nicht beliebig, sondern basieren auf den im ersten Hauptteil vorgestellten Autoadaptionprozessen und ermöglichen dem Nutzer durch eine Interpretation die Entdeckung beziehungsweise Erfindung von Neuem.

Nicht die endlose Unübersehbarkeit dessen, was noch nicht gesehen ist, »bewirkt« die Neugier [...]. Auch wenn man alles gesehen hat, dann erfindet gerade die Neugier Neues.

(Heidegger 1927, S. 348)

Neugierige MLA erwecken somit den Anschein, menschliche Neugier durch ein **als-freies** Erfassen technisch unterstützen zu können. Solch eine Perspektive ist besonders dann interessant, wenn angenommen wird, dass reine Neugier im Sinne eines als-freien Erfassens für das Dasein zumindest einer gewissen Umstellung bedarf oder gar prinzipiell unmöglich ist.

Das schlichte Sehen der nächsten Dinge im Zutunhaben mit... trägt die Auslegungsstruktur so ursprünglich in sich, daß gerade ein gleichsam als-freies Erfassen von etwas einer gewissen Umstellung bedarf.

(Heidegger 1927, S. 149)

Neugierige MLA erzeugen Strukturvorschläge, die un-vorher-gesehene Strukturen präsentieren, da sie Rohdaten aufnehmen und dabei die impliziten Konzepte beziehungsweise Vorannahmen der Nutzer ignorieren können. Ein Beispiel Heideggers für solch eine implizite Vorannahme betrifft das Hören einer Fremdsprache.

Sogar dort, wo das Sprechen undeutlich oder gar die Sprache fremd ist, hören wir zunächst unverständliche Worte und nicht eine Mannigfaltigkeit von Tondaten.

(Heidegger 1927, S. 164)

¹¹⁰ Bezüglich des Kontinuums zwischen zielorientierten und neugierigen Artefakten ist festzuhalten, dass bei MLA immer der Begriff der Instanziierung mitgedacht werden muss und innerhalb der Skala diesbezüglich kein Bruch auftritt.

MLA können auch die Interpretation von Eingabedaten als eine Mannigfaltigkeit von Tondaten vermeiden und völlig ohne Vorstrukturierung mit den Rohdaten arbeiten. Künstliche neuronale Netze sind etwa aufgrund genau dieser Idee gerade nicht an menschliche Gehirne angelehnt und wesentlich weniger stark vorstrukturiert (Teuchert-Noodt 2011).

Der obigen Verwendung der Neugier als Problembegriff steht zunächst noch Heideggers Bezug auf das „zunächst Zuhandene“ im Weg, aus dem in die ferne Welt tendiert wird¹¹¹. Hier scheint sich eine Analogie dazu zu finden, dass innerhalb der Informatik eine Bezeichnung der Rohdaten als Instanzen üblich ist, die impliziert, dass die Daten eine bereits bekannte Struktur instanziierten. Diese Redeweise wurde im Vorherigen bereits abgelehnt, da Strukturvorschläge auch und gerade dann eingesetzt werden, wenn die Auffindbarkeit von relevanten Strukturen in den Rohdaten lediglich vermutet oder erwünscht ist. Insbesondere wenn einem MLA eine Lernstrategie ohne explizite Ziele zugrunde liegt, etwa **selbstorganisierende Karten**, kann kein Anschein einer Zuhandenheit der Eingabedaten festgestellt werden¹¹². Es ist nicht sofort klar, ob der Bezug Heideggers auf das Zuhandene ein Hindernis für die Verwendung von Neugier als Problembegriff darstellt, denn bei Heidegger wird durch diese Formulierung vor allem betont, dass das Dasein immer schon in Welt ist und sich nicht im Modus der Neugier erst Welt sucht oder erschafft. Unabhängig davon scheint der Bezug auf Zuhandenes die Diskussion maschinellen Lernens zu verunklaren und nicht konstruktiv zu irritieren.

Als Veranschaulichung der Frage, was genau Eingabedaten in der Praxis instanziierten, soll ein maschinell lernendes Schachprogramm betrachtet werden, das auf der Vermutung basiert, dass es eine Gewinnstrategie oder zumindest eine Remisstrategie für Weiß gibt. Als Aufgabe des Artefaktes wird die Suche nach ebenjener Strategie festgelegt. Das gestiftete Worumwillen ist dementsprechend der Sieg in einem Schachspiel¹¹³. Die abgeschlossenen Schachspiele, die als Eingabedaten verwendet werden, sind hier keine Instanzen im eigentlichen Sinn. Die Eingabedaten werden zwar im Rahmen des Autoadaptionsprozesses des MLA verwendet, aber es ist für keine der Parteien im Vorhinein bekannt, ob eine gute oder schlechte Strategie zum Sieg geführt hat, beziehungsweise welche Teile der Partie Ausdruck einer guten oder schlechten Strategie sind. In der Praxis der Erstellung eines Schachprogramms äußert sich diese Schwierigkeit insbesondere im Versuch, das Programm Fehler machen zu lassen.

¹¹¹ Die Frage, ob nach Beseitigung dieses Hindernisses die Rede von neugierigen MLA eine stabile Analogie zwischen Heideggers Daseinsanalytik und dem maschinellen Lernen darstellt, muss selbstverständlich verneint werden – etwa müsste hierzu begründet werden, inwiefern es eine Rolle spielt, dass die Neugier gerade die uneigentliche Form des Verstehens darstellt. Das Interesse dieser zweckrationalen Nutzung der Daseinsanalytik beschränkt sich jedoch darauf, die Neugier und die Vor-Struktur als Begriffe zu gewinnen, die eine konstruktive Irritation erzeugen und dabei irreführende Assoziationen aus dem Weg zu räumen und hilfreiche Ideen zu bestärken. Der Bezug Heideggers auf ein Zuhandenes birgt die Gefahr solch einer irreführenden Assoziation und wird aus diesem Grund gesondert diskutiert.

¹¹² Im Gegensatz dazu scheinen übergebene Rohdaten genau dann einem gegebenen Bestand zu entsprechen, wenn sie als tatsächlich instanziiierende Instanzen auftreten und im Rahmen des Autoadaptionsprozesses als solche eingesetzt werden. Das heißt, ein Eingabedatum kann nur zuhanden erscheinen, wenn dem betrachteten MLA ein hohes Maß an Vorwissen und Konzepte mitgegeben wurden. In diesem Fall ist der Möglichkeitsraum der Daten thematisch genau erfasst und es werden stark vorstrukturierte MLA eingesetzt, die so konstruiert wurden, dass sie die den Möglichkeitsraum bereits vor Berücksichtigung der Eingabedaten möglichst gut abbilden.

¹¹³ Hierbei handelt es sich genau dann um einen Zweck, wenn der Sieg des Artefaktes dem Nutzer bereits ausreicht. Im Weiteren wird jedoch davon ausgegangen, dass der Nutzer seine eigene Spielstärke verbessern möchte.

A PC program at its maximum strength will wipe out any casual player without mercy. Ironically, the main task of chess software companies today is to find ways to make the program weaker, not stronger, and to provide enough options that any user can pick from different levels and the machine will try to make enough mistakes to give him a chance.

(Kasparov 2010)

Die Fähigkeit, auf Basis der Eingabedaten bereits bekannte Stellungen als solche zu identifizieren und positiv oder negativ zu bewerten, ist zwar ein zentraler Bestandteil im Schachspiel, die eigentliche Strategie besteht jedoch darin, die positiv bewertete Stellung mit der eigenen Farbe auch zu erreichen und nicht lediglich deren Nützlichkeit zu konstatieren. Die Eingabedaten stellen aus dieser Perspektive Instanzen für Schachspiele und nicht für Siegstrategien dar. Die Nichttrivialität des Schrittes von der Beobachtung eines Schachspiel zur Formalisierung einer Gewinnstrategie ist meist genau der Grund, aus dem Schachprogramme überhaupt erstellt wurden. Erst diese Fähigkeit erlaubt es dem Nutzer, ein Schachspiel als eine Instanz einer Strategie zu erkennen beziehungsweise zu verstehen.

Das Problem bei der Rede von neugierigen MLA besteht nun darin, dass neugierige MLA kein Konzept von Unverwendbarkeit zu haben scheinen. Sie adaptieren ihre Struktur oder den entstehenden Strukturvorschlag in Reaktion auf *alle* fremden und fernen Rohdaten. Die Autoadaption der neugierigen MLA erweckt entsprechend nicht den Anschein, anhand der Aufdringlichkeit, Auffälligkeit oder Aufsässigkeit der Eingabedaten zu erfolgen, sondern geschieht in Reaktion auf *jede* Eingabe. Speziell für zielorientierte MLA entsteht dementsprechend nicht der Anschein, dass sinnvoll von Zuhandenheit und damit von Vorhandenheit im Sinne einer Unzuhandenheit gesprochen werden kann¹¹⁴. Es müsste in Folge eine Art von **Abhandenheit** der Eingabedaten als Teil der fernen und fremden Welt konstatiert werden, um die konstruktive Irritation der Rede von Neugier aufrecht zu erhalten. Neugierige MLA passen die von ihnen erstellten Strukturvorschläge dieser fernen und fremden Welt an und scheinen dabei weg von den bisherigen Strukturvorschlägen zu tendieren. Der Autoadaptionsprozess besteht darin, dass ein entstehender Strukturvorschlag fortwährend auf Basis neuer sensorischer Eindrücke aktualisiert wird.

Die freigewordene Neugier besorgt aber zu sehen, nicht um das Gesehene zu verstehen, das heißt in ein Sein zu ihm zu kommen, sondern nur um zu sehen. Sie sucht das Neue nur, um von ihm erneut zu Neuem abzuspringen.

(Heidegger 1927, S. 172)

Insgesamt scheint sich die Reaktion nicht vorstrukturierter MLA auf kontingent erscheinende Sensordaten sehr gut als die fortwährend zu Neuem abspringende Suche eines **neugierigen maschinell lernenden Artefaktes** beschreiben zu lassen.

¹¹⁴ Zielorientierte MLA hingegen scheinen, wie bereits beschrieben, durchaus ein Konzept von Unverwendbarkeit zu besitzen und je nach Form des Autoadaptionsprozesses selbsttätig und unabhängig vom Menschen ihre ursprüngliche eigene Struktur, das gestiftete Ziel oder beides zu verdecken und zu modifizieren, um die Aufdringlichkeit, Auffälligkeit oder Aufsässigkeit der Eingabedaten zu überwinden.

Vor-Struktur als Problembegriff

Im Vorherigen wurde der Entwurf als ungeeignet zur Erzeugung einer konstruktiven Irritation zurückgewiesen während die Neugier als Problembegriff freigelegt werden konnte. In diesem letzten Schritt der Suche nach Problembegriffen wird anknüpfend an die Betrachtung des Entwurfs über den Begriff der Auslegung die Vor-Struktur als ein zweiter hilfreicher Problembegriff freigelegt werden. Die Unterscheidung von Vor-Strukturen wird als eine Möglichkeit identifiziert werden, detailliert zwischen einzelnen Formen maschinellen Lernens zu unterscheiden und insbesondere klarer benennen zu können, in welcher Hinsicht die MLA oder die zugrunde liegenden Lernstrategien selbsttätig agieren.

Dieser Schritt der Suche beginnt zunächst mit der Betrachtung der Auslegung, die in der Daseinsanalytik die Ausbildung des Verstehens bezeichnet.

Das Entwerfen des Verstehens hat die eigene Möglichkeit, sich auszubilden. Die Ausbildung des Verstehens nennen wir Auslegung.

[...]

Die Auslegung ist nicht die Kenntnisaufnahme des Verstandenen, sondern die Ausarbeitung der im Verstehen entworfenen Möglichkeiten.

(Heidegger 1927, S. 148)

Die konstruktive Irritation der quasi-intentionalen Rede in Bezug auf die Auslegung tritt auf, wenn davon gesprochen wird, dass ein MLA im Rahmen des Autoadaptionsprozesses in die Ausarbeitung der Möglichkeiten zur Autoadaptation zu streben scheint. Diese Formulierung impliziert, dass es wirkt, als ob ein MLA sowohl die Roh- beziehungsweise Trainingseingaben als auch seine eigene Parametrisierung auslegt. Das wiederum bedeutet, dass in der Rede von Auslegung das Phänomen, dass MLA ihre eigene Struktur modifizieren, beobachtet werden kann. Dies ist ein sehr interessanter Punkt, da ein Autoadaptionsprozess je nach Lernstrategie sehr unterschiedliche Vorgaben erhalten kann, die es ihm erlauben, auf sehr verschiedene Weisen seine eigene Struktur zu modifizieren. Diese Vielfalt an Optionen zur Modifizierung der Struktur eines MLA soll im Weiteren gezielt im Fokus stehen. Diese Verschiebung des Fokus¹¹⁵ identifiziert den Begriff der **Vor-Struktur**, in der die Auslegung gründet, als einen im Kontext des maschinellen Lernens sehr interessanten Problembegriff¹¹⁵. Das Alleinstellungsmerkmal von MLA – der Autoadaptionsprozess – besteht gerade darin, dass MLA neue Rohdaten verarbeiten und dabei den Anschein erwecken, ihre Vor-Struktur zu modifizieren, das heißt, ihre Struktur beziehungsweise die bereits verarbeiteten Daten auszulegen. Im Rahmen des Autoadaptionsprozesses verdecken und modifizieren sowohl zielorientierte als auch neugierige MLA ihre ursprüngliche Vor-Struktur – je nach Lernstrategie in sehr unterschiedlichem Ausmaß und auf sehr unterschiedliche Weise. Inwieweit diese Modifikationen selbsttätig, das heißt unabhängig vom Menschen, durchgeführt werden, unterscheidet sich stark zwischen den einzelnen Lernstrategien und hängt nicht direkt mit der Unterscheidung in zielorientierte und neugierige MLA zusammen.

¹¹⁵ Der Begriff der Vor-Struktur bezeichnet hier die Gesamtheit der über den Strukturvorschlag hinaus bestehenden Strukturen, die die Reaktion eines MLA auf Eingabedaten festlegen und damit den Autoadaptionsprozess strukturieren.

Auch wenn der Nutzer zu Beginn des Autoadaptionsprozesses die Vor-Struktur noch verstanden hat oder zumindest kannte, kann sich das im Laufe des Autoadaptionsprozesses ändern. Natürlich ermöglichen auch nicht maschinell lernende Artefakte mitunter das vorprädikative Sehen des Zuhandenen und beeinflussen damit die Wahrnehmung, allerdings operieren diese Arten von Artefakten unter fixen und meist transparenten Rahmenbedingungen. MLA unterscheiden sich von diesen Artefakten insbesondere durch ihre meist opake – mitunter sogar autoadaptive – Vor-Struktur und die Selbsttätigkeit bei der Modifikation beziehungsweise Generierung dieser Vor-Struktur. Die möglichen Eigenschaften der Vor-Struktur werden in Abschnitt 3.5.3 im Rahmen der Diskussion nichttrivialer Technik genauer betrachtet, es sei jedoch schon festgehalten, dass die Vor-Struktur und ihr Agieren zwar opak erscheinen können, aber dadurch nicht unsichtbar werden und es auch nicht zwingend zu einem Verlust der **Spuren** (Hubig 2008, S. 183ff) kommt.

Analogien zur Vor-Struktur des Daseins sind an dieser Stelle zwar möglich¹¹⁶, aber es ist zum besseren Verständnis maschinellen Lernens nicht notwendig, der Begriffsbildung Heideggers weiter zu folgen als bis zu der Feststellung, dass eine Vor-Struktur besteht und dass diese autoadaptiv und somit variabel ist. Die Vor-Struktur jeder Lernstrategie kann und sollte einzeln diskutiert werden und dies findet seitens der Informatik mitunter auch statt. Dort werden beispielsweise unterschiedliche Konzepte für die Identifikation systematischer Fehler von MLA diskutiert und es wird versucht, diese bei der Entwicklung von Autoadaptionsprozessen zu vermeiden oder zumindest transparent zu machen. Bezüglich der gezielten Nichtbetrachtung einer engen Analogie zur Vor-Struktur des Daseins ist anzumerken, dass Heidegger selbst davon ausgeht, dass der Begriff der Auslegung zumindest im Rahmen der Untersuchung des Daseins nur Verwendung finden darf, wenn alle drei Teile der Vor-Struktur des Daseins auftreten. Er führt dies am Begriff der Aussage ein, den er als einen Modus von Auslegung identifiziert.

Mit welchem Recht fassen wir überhaupt die Aussage als Modus von Auslegung? Ist sie so etwas, dann müssen in ihr die wesenhaften Strukturen der Auslegung wiederkehren.

[...]

Die Aussage hat notwendig wie Auslegung überhaupt die existenzialen Fundamente in Vorhabe, Vorsicht und Vorgriff.

(Heidegger 1927, S. 156f)

Ein gestiftetes MLA begreift Eingabedaten zwar auf Basis einer Vor-Struktur, die zum Einsatz kommende Vor-Struktur kann jedoch für verschiedene MLA sehr unterschiedlich aussehen. In jedem Fall unterscheidet sie sich stark von der Vor-Struktur des Daseins, dem Tripel aus **Vorhabe**, **Vorsicht** und **Vorgriff**.

3.4.3. Neugier und Vor-Struktur jenseits von Heidegger

Die übergeordnete Absicht des Abschnitts 3.4 bestand im Versuch, präziser zu fassen, auf welchen Aspekten des maschinellen Lernens die eingangs beschriebene quasi-intentionale

¹¹⁶ Eine derartige Analogie entwickelt Wei-Ding (Wei-Ding 2011, S.85).

Redeweise basieren kann. Die zentralen Aspekte des In-der-Welt-seins wurden zu diesem Zweck auf ihre Nützlichkeit zur Lokalisierung des aus der quasi-intentionalen Redeweise resultierenden Unbehagens hin untersucht. Mit der Vor-Struktur und der Neugier wurden zwei Quellen des Unwohlseins identifiziert und als konstruktive Problembegriffe isoliert. Im Weiteren wird nun zusammengefasst, welchen Mehrwert die Auswahl der Problembegriffe unabhängig von Heidegger in der nachfolgenden Diskussion des maschinellen Lernens bietet.

Das Entstehen von Assoziationen ist sowohl eine Herausforderung als auch das zentrale Potenzial der hier vorgenommenen Suche nach Problembegriffen. Die Schwierigkeit liegt in der grundsätzlichen Frage, wie sinnvoll mit Begriffen gearbeitet werden kann, die eine Vergleichsperspektive herausfordern und dadurch auch Unangemessenes in die Rekonstruktion einbringen, wodurch sie drohen zu verschleifen, was am maschinellen Lernen das Spezifische zu sein scheint. Eine Antwort liegt darin, dass nicht einfach das Heidegger'sche Konzept des In-der-Welt-seins im Kontext des maschinellen Lernens wieder auftaucht. Die implizite Unterstellung, KNN wären eine Form von Dasein, wird nicht unternommen. Der Mehrwert liegt stattdessen in der Betrachtung der *spezifischen* Differenzen des In-der-Welt-seins in der Daseinsanalytik auf der einen Seite und der Rolle, die einzelne Momente des Begriffes im Kontext nicht vorstrukturierter KNN übernehmen können, auf der anderen Seite. Die Diskussion des maschinellen Lernens kann auf Bereiche fokussiert werden, die zwar immer noch eine Ungeklärtheit aufweisen, jedoch weniger pauschal und besser handhabbar sind als zuvor. Die gezielt nicht verdeckten, spezifischen Differenzen bestehen etwa in der unterschiedlichen Konnotation von Neugier. Die Neugier wird im Kontext des maschinellen Lernens positiv dargestellt – bei Heidegger hingegen ist Neugier über den Zusammenhang mit Gerede und Verfallen negativ belegt. Die Feststellung, MLA eine Vielzahl von Eigenschaften und Momenten *nicht* sinnvoll zusprechen zu können, ist in sich bereits ein Mehrwert. Die Identifizierung zielorientierter MLA als eines Teilbereichs des maschinellen Lernens, der die beschriebene Irritation *nicht* auslöst, stellt etwa einen Schritt in genau diese Richtung dar. Dies schließt den Einsatz anderer Problembegriffe nicht aus, aber zumindest *ein* besonders prägnanter Ursprung der ursprünglichen Irritation ist jetzt präziser lokalisierbar als zuvor.

Die vorangegangene Suche nach Problembegriffen diene nicht nur der Vorbereitung einer technikphilosophischen Diskussion des maschinellen Lernens. Die Ergebnisse bieten darüber hinaus bereits erste Möglichkeiten zur Bewertung des Ausmaßes der Selbstorganisation von MLA. Die Identifizierung eines MLA als neugierig oder zielorientiert und die Einschätzung der Flexibilität seiner Vor-Struktur können als Ansätze betrachtet werden, unterschiedliche Aspekte von Selbstorganisation besser unterscheid- und greifbar machen. Diese Aspekte sind nicht nur aus Sicht der Technikphilosophie relevant, sondern insbesondere auch für interdisziplinäre Diskussionen unmittelbar nutzbar. Beispielsweise kann und wird auch aus anderen Perspektiven als denjenigen der Medienphilosophie oder Phänomenologie diskutiert, ob und inwiefern Welt technisch vermittelt sein kann. Ein prominenter Teilnehmer dieser Diskussion entstammt der Physik¹¹⁷.

¹¹⁷ Tatsächlich entstammt der Kommentar sogar einer Diskussion über die **Quantenphysik**, die später im Ausblick noch einmal aufgegriffen wird.

Whether you can observe a thing or not depends on the theory which you use. It is the theory which decides what can be observed.

(Einstein 1926)

Die Möglichkeit so zu argumentieren, ohne auf die Daseinsanalytik oder andere Theorien zurückgreifen zu müssen, liefert ein starkes Argument für die interdisziplinäre Nutzbarkeit der obigen Ergebnisse. Auch wenn die Herleitung und Motivation der Rede von Neugier im Kontext des maschinellen Lernens keine Beachtung findet – was aus Sicht der Informatik zwangsweise der Fall ist – besitzt die Idee, „Neugier“ und „Vor-Struktur“ als Konzepte zur Unterscheidung verschiedener Teilbereiche des maschinellen Lernens beziehungsweise unterschiedlicher Autoadaptionsprozesse einzusetzen, noch genügend Tragkraft. Das Gleiche gilt für die Beschreibung von Strukturvorschlägen neugieriger MLA als un-vorher-gesehen. Diese Bezeichnungen sind intuitiv verständlich, erzeugen hilfreiche Assoziationen und stellen damit eine deutlich präzisere Formulierungen dar als etwa die Rede von überraschenden Strukturvorschlägen. Insgesamt bietet es sich dementsprechend an, einen zugänglicheren und weniger voraussetzungsreichen Neugierbegriff als denjenigen Heideggers zu entwerfen, der dennoch die hilfreichen Facetten und Intuitionen der obigen Diskussion aufrechterhält. Das Konstrukt einer solchen Form von Neugier kann sich an einer traditionellen, positiven Konnotation von Neugier orientieren – ob darüber hinaus die Kopplung an die Idee der Vermittlung beziehungsweise Induktion eines Weltbezuges notwendig mitgedacht werden müsste, ist zunächst noch eine offene Frage und soll als Anknüpfungspunkt an die vorliegende Arbeit festgehalten werden. Einstweilen fehlen noch begriffliche Details, aber die schon erzielten Gewinne sind bereits spürbar, auch und insbesondere wenn Neugier und Vor-Struktur „nur“ als Überschriften von Problembereichen verstanden werden – und genau das wird von einer Eröffnung des Themengebietes für die technikphilosophische Diskussion verlangt.

Auch wenn die Entwicklung eines zugänglicheren Begriffes von Neugier noch aussteht, fand mit der Setzung des zielorientierten und neugierigen maschinellen Lernens bereits eine vergleichsweise konkrete Begriffsbildung statt¹¹⁸. Diese Unterscheidung zweier Formen maschinellen Lernens wird nachfolgend noch einmal explizit zusammengefasst: **Zielorientiertes maschinellen Lernen** weist ein Ziel auf, von dem bestimmbar ist, wann es erreicht wurde und das meist in Form einer Zielfunktion auftritt. Die Nutzer dieser Artefakte haben ein klar formulierbares Problem und erwarten von dem eingesetzten MLA ein bestimmtes Ergebnis. Abstrakte Beispiele für zielorientierte Autoadaptionsprozesse sind Optimierungsprobleme, durchaus auch im mathematischen Sinn. Im Rahmen von Optimierungsproblemen sollen Vorgehensweisen effizient gestaltet oder Ausgabewerte maximiert werden. Diese Optimierungen können durchaus in Form maschinellen Lernens auftreten, etwa wenn das MLA bestimmt, welche zusätzlichen Messungen durchgeführt werden sollten, um die Qualität des Strukturvorschlages in Bezug auf die Zielfunktion zu bewerten. Allerdings können die Messwerte auch ohne den Einsatz maschinellen Lernens ausgewertet werden und der Übergang zu Optimierungsalgorithmen, die keine Eingabedaten erhalten, sondern interne mathematische Funktionen auswerten, ist für den Laien mitunter schwierig zu erkennen und abhängig von der Definition der Systemgrenzen. **Neugieriges**

¹¹⁸ Die entsprechenden Überlegungen zum Begriff der Vor-Struktur werden im Abschnitt über Nichttrivialität (3.5.3) folgen.


maschinelles Lernen hingegen umfasst diejenigen MLA, die zumindest den Anschein erwecken, ohne Zielvorgaben initiiert worden zu sein. In diesem Fall wird vom Nutzer kein Vorwissen investiert, das über sehr allgemeine Konzepte zur Vorgehensweise hinausgeht, wie es etwa eine Vorgabe der Eingabereihenfolge der Rohdaten darstellt. Auch im neugierigen Lernen können im Rahmen einer Codierung zulässige Aktionen definiert und somit Strukturvorgaben gemacht werden.

Die Einordnung von Autoadaptionsprozessen auf dem Kontinuum von neugierigen bis zielorientierten MLA lässt sich nicht pauschal an der Wahl der eingesetzten Lernstrategien festmachen¹¹⁹. Allerdings bietet sich der Einsatz von stark mathematisch geprägten Konzepten wie den **Stützvektormethoden** und dem statistischen Lernen besonders dann an, wenn die Aufgabenstellung einem Optimierungsproblem ähnelt. Gerade **statistisches Lernen** bietet zwar durchaus die Möglichkeit, ziellose Autoadaptionsprozesse zu initiieren, aber die mathematische Theorie und die heuristischen Kenntnisse der Nutzer sind so stark ausgeprägt, dass nur in Ausnahmefällen unvorhergesehene Strukturvorschläge entstehen oder überhaupt angestrebt werden. Unvorhergesehene Ergebnisse deuten in solchen, mathematisch anspruchsvollen, Codierungen häufig schlicht auf nutzerseitige Fehler beim Einsatz des MLA hin. Aus solchen Fehlern könnte prinzipiell ein durch den Nutzer interpretierbarer Strukturvorschlag gewonnen werden, allerdings wird in der Praxis im Vorfeld einer mathematischen Optimierung eine vergleichsweise klare **Erwartung** an den Strukturvorschlag bestehen und der unvorhergesehene Vorschlag wird allenfalls eine Irritation beim Nutzer auslösen, bevor er verworfen wird. **Evolutionäres Lernen**, **KNN** und **instanzenbasiertes Lernen** hingegen werden durchaus für die Implementierung zielloser Autoadaptionsprozesse eingesetzt.¹²⁰ Diese drei Lernstrategien realisieren neugierige Autoadaptionsprozesse dabei auf sehr unterschiedliche Weisen. Die Vorgabe einer Fitnessfunktion im evolutionären Lernen stellt etwa eine der Strukturvorgaben dar, die zwar mit neugierigem Lernen vereinbar ist, aber genauso auch zu einem zielorientierten Autoadaptionsprozess führen kann.

Insgesamt kann festgehalten werden, dass die Analyse der als neugierig bezeichneten maschinell lernenden Artefakte technikphilosophisches Potenzial bietet, da hier ein Technikbereich identifiziert wurde, dessen Realisierungen bestehende Technik und sich selbst systematisch verändern können. Im Gegensatz dazu unterscheidet sich der Teilbereich maschinellen Lernens, der als zielorientiertes Lernen bezeichnet wurde, technikphilosophisch nicht sehr stark von nicht-lernenden Algorithmen. Diese Feststellung erlaubt eine deutliche Präzisierung der Diskussion maschinellen Lernens, da nicht mehr pauschal auf alle MLA referenziert werden muss. Die Unterscheidung von Vor-Strukturen wird erst im Rahmen der Diskussion der Nichttrivialität in Abschnitt 3.5.3 genauer untersucht. Dieser Problembegriff wird sich als hilfreich erweisen, wenn das Ausmaß der Autoadaptivität eines MLA untersucht werden soll. Zusammengefasst markieren die Begriffe der Neugier und der Vor-Struktur somit Fortschritte bei der Trennung derjenigen Teilbereiche des maschinellen Lernens, die einer genaueren technikphilosophischen Betrachtung bedürfen und derjenigen Teile, bei denen

¹¹⁹ Weiterhin entspricht die Begriffsbildung nicht einfach der Unterscheidung subsymbolischer und symbolischer Autoadaptionsprozesse oder der Entscheidung, gewisse MLA als Black Box zu betrachten.

¹²⁰ Auch diese Formen maschinellen Lernens werden in der Praxis selbstverständlich zum größten Teil zur Lösung von klar umrissenen und intensiv vorbereiteten Problemstellungen eingesetzt, da den Nutzern normalerweise an der performanten Lösung eines konkreten Problems gelegen ist. Im Falle des Einsatzes von evolutionärem Lernen zur Optimierung eines Flugzeugmodells in Abschnitt 3.2.2 etwa ist zwar der Kontext des Problems sehr komplex, aber es wird auch dort einiger Aufwand betrieben, um das evolutionär lernende MLA mit einer möglichst passenden Vorstruktur zu versehen.



zumindest mit Begriffen wie Selbstorganisation nicht sinnvoll argumentiert werden kann oder sollte.

3.5. Abgrenzung von aktuellen technikphilosophischen Entwürfen

Der erste Schritt der Analyse maschinellen Lernens bestand in der Darstellung der Perspektive der Informatik und der impliziten Feststellung, dass die Informatik sich selbst bezüglich der von ihr eingesetzten Technikkonzepte tendenziell eine Black Box ist¹²¹. In der Argumentation des zweiten Hauptteils wurde daran anknüpfend eine Abgrenzung von etablierten Begriffsbildungen vorgenommen und die prinzipielle Verortung von MLA als Technik problematisiert. Anschließend wurden mit Hilfe von Heideggers Modellierung des menschlichen Zugangs zur Welt mögliche Problembegriffe freigelegt. Die Absicht hinter der Betrachtung der Modellierung Heideggers lag darin, konkretere Fragerichtungen zu identifizieren, die in Hinblick auf maschinelles Lernen einen Erkenntnisgewinn liefern können. Vor der darauf aufbauenden Betrachtung einer zweiten Modellbildung – dem Entwurf Goodmans in *Ways of Worldmaking* – und einer techniknahen Aufarbeitung von Goodmans Analysen in Hinblick auf maschinelles Lernen soll zunächst geklärt werden, ob und in welcher Hinsicht eine solche Aufarbeitung bereits stattgefunden hat oder als noch ausstehend bezeichnet werden kann.

Zu diesem Zweck soll im Folgenden dargestellt werden, welche, das Gebiet des maschinellen Lernens betreffenden, Fragen und Antworten bereits in aktuellen technikphilosophischen Entwürfen formuliert wurden. Das Ziel wird dabei sein, zu prüfen, welches der Konzepte im Kontext des maschinellen Lernens zum Einsatz kommen kann und ob gegebenenfalls sogar bereits eine Unterscheidung zwischen neugierigem und zielorientiertem maschinellem Lernen möglich ist. Die betrachteten technikphilosophischen Entwürfe müssen dementsprechend eine thematische Nähe zur Informatik aufweisen. Diskussionen, die den Bereich der **konvergierenden Technologien** und damit die interdisziplinären Betrachtungen von **NBIC-Technik**¹²² zumindest mit im Blick haben, erfüllen diese Anforderung. Weiter scheint in diesen Bereichen – unabhängig von den Herausforderungen, die sich aus der These einer Konvergenz verschiedener Technikbereiche ergeben – Technik entstanden zu sein und neu zu entstehen, die neuartige Begriffe von Technik erfordert (Roco et al. 2003). Mit den Konzepten der naturalisierten Technik (Nordmann 2008, S. 173) und der transklassischen Technik (Hubig 2008, S. 165) liegen zwei Perspektiven vor, mit Hilfe derer moderne und insbesondere in der Informatik beheimatete Hochtechnik diskutiert werden kann. Im Folgenden sollen diese beiden Perspektiven in Bezug auf maschinell lernende Artefakte zum Einsatz gebracht werden. Das Ziel dieses Abschnittes wird zunächst darin liegen, die Schwerpunkte beider Perspektiven vergleichbar darzustellen und anschließend zu betrachten, wie die Diskussion maschinellen Lernens von den beiden Perspektiven profitieren kann. Gleichzeitig sollen die Grenzen, beziehungsweise die sich unterscheidenden Interessenschwerpunkte, aufgezeigt werden. Als dritte technikphilosophische Perspektive wird anschließend das Konzept der nichttrivialen Maschine analysiert. Hier liegt bereits eine Analyse von Kaminski vor, die direkt auf den Kontext des maschinellen Lernens eingeht und auf der aufgebaut werden kann (Kaminski 2012).

¹²¹ Insbesondere aufgrund des gewollten Fokus' auf Performanz und andere Leistungskennzahlen.

¹²² Diese umfasst die Bereiche der Nanotechnologie, Biotechnologie, Informationstechnologie und der Neurowissenschaften. Eine Darstellung des Diskurses bietet Kogge (Kogge 2008, S. 940ff).

3.5.1. MLA als naturalisierte Technik

Zunächst sollen die Begriffe und Kriterien der **naturalisierten Technik** dargestellt und auf das maschinelle Lernen übertragen, sowie die Potenziale und Grenzen dieser Übertragung identifiziert werden. Im ersten Schritt dieser Analyse werden die Kriterien für die Verortung von Artefakten als naturalisierte Technik so umformuliert, dass sie vergleichbar zu denjenigen der transklassischen Technik werden. Zu diesem Zweck werden zunächst die zwei von Nordmann genannten Kriterien mit Hilfe eines Mengendiagramms veranschaulicht (Visualisierung A). Diese Darstellung wird anschließend über zwei Zwischenschritte in eine Zusammenstellung von drei Kriterien umgeformt (Visualisierung D), und es wird argumentiert, inwiefern die neue Bestimmung sowohl präziser als auch von größerem Nutzen in der weiteren Diskussion ist. Die beiden Zwischenschritte der Argumentation werden ebenfalls als Mengendiagramme veranschaulicht (Visualisierungen B und C). Die resultierende Zusammenstellung von drei Kriterien ist zwar gezielt so gewählt, dass sie sich auch für die Betrachtung der transklassischen Technik einsetzen lässt, sie ergibt sich jedoch direkt aus der Arbeit mit dem Begriff der naturalisierten Technik.

Naturalisierte Technik erscheint nicht nur deshalb vielversprechend, weil sie explizit die Informatik in ihre Betrachtungen einbezieht, sondern auch weil Nordmann seine Begriffsbildung sogar mit Bezug auf Selbstorganisation motiviert.

Instead of signifying transparency, rationalization, and control, “technology” becomes opaque, magical, even uncanny. [...] This limit could also be reached where engineering seeks to exploit surprising properties that arise from natural processes of self-organization.

(Nordmann 2008, S. 175)

Diese Beschreibung passt auch auf einige maschinelle Lernstrategien und die Rede von der Nutzung überraschender oder unerwarteter Ergebnisse von Selbstorganisationsprozessen scheint sich gut auf maschinelles Lernen übertragen zu lassen. Maschinell lernende Artefakte wären möglicherweise Technik genau auf der von Nordmann angedeuteten Grenze zwischen naturalisierter und nicht-naturalisierter Technik. Die beiden wichtigsten Kriterien zur Beurteilung, ob diese Grenze überschritten ist, sind für Nordmann Wahrnehmbarkeit und Kontrollierbarkeit.

The hallmark of technology naturalized is that it acts below or above the thresholds of perception and control, that we cannot represent its agency as it occurs, that we have no switches to initiate or stop operation, no direct knowledge of whether it is functioning or broken down.

(Nordmann 2008, S. 177)

Diese beiden Kriterien sollen anhand der Beispiele veranschaulicht werden, die Nordmann selbst nennt, ergänzt um Fälle, die nur eines der beiden Kriterien erfüllen.

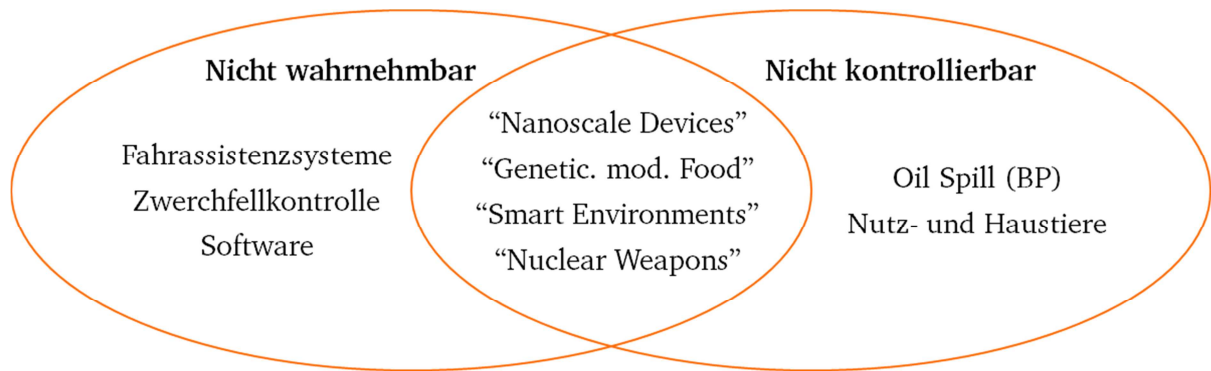


Abbildung 40: Beispiele für naturalisierte Technik

Der Begriff der Kontrollierbarkeit ist hier nicht eindeutig bestimmt und die Beispiele sollen in erster Linie der Illustration dienen, da der Begriff im Weiteren noch ersetzt wird. In diesen Beispielen deuten sich dennoch bereits die Grenzen der Verortbarkeit von maschinell lernenden Artefakten der Autoadaptionsprozess meist ohne Probleme aufzuhalten ist. Vor den genaueren Betrachtungen des maschinellen Lernens wird im Folgenden die Charakterisierung naturalisierter Technik präzisiert und mit transklassischer Technik vergleichbar dargestellt. So wird herausgearbeitet, was das „Neue“ an naturalisierter Technik ist und wie maschinell lernende Artefakte sich dazu verhalten. Die Einordnung von Technik oder Lernstrategien im Einzelfall soll hingegen nicht diskutiert werden¹²³.

Präzisierung der Visualisierung naturalisierter Technik

Im ersten Schritt wird eine Visualisierung erstellt, die den Raum der naturalisierten Technik mit Hilfe positiver Kriterien darstellen wird. Hierzu wird der Raum der möglicherweise unbeschränkten Technik als Betrachtungsrahmen festgehalten.

Human engineering is thus adding further causal processes that operate behind our backs and may or may not produce catastrophic consequences of their own.

(Nordmann 2008, S. 181)

Innerhalb des Raums der möglicherweise unbeschränkten Technik werden die Räume der wahrnehmbaren beziehungsweise kontrollierbaren Technik durch ovale Flächen visualisiert. Entsprechend kann naturalisierte Technik nur im orange dargestellten Außenbereich aufgefunden werden.

¹²³ Dies wäre beispielsweise die Frage, ob maschinell lernende Artefakte des **Pervasive Computing**, die auf Gesten des Nutzers reagieren, ohne dass sich der Nutzer dessen notwendigerweise bewusst ist, dennoch als durch den Nutzer kontrolliert betrachtet werden müssen oder ob in diesem Fall eher ein Nutzerstereotyp als der konkrete Nutzer die Technik kontrolliert.

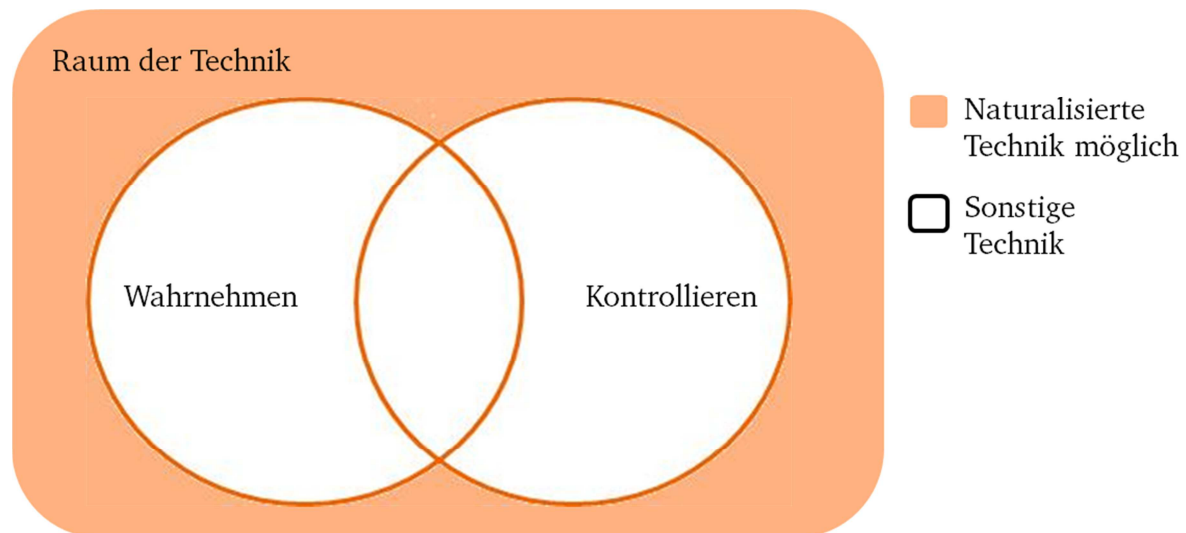


Abbildung 41: Raum naturalisierter Technik – Visualisierung A

Diese Darstellung soll die Intuition vermitteln, dass es sich bei naturalisierter Technik um etwas Außergewöhnliches handelt, das außerhalb der Reichweite der Nutzer liegt.

More briefly put, we encounter naturalized technology when, for all we know, a technical agency unfolds below or above human thresholds of perception and control.

(Nordmann 2008, S. 176)

Wie bereits angedeutet ist der Begriff der Kontrollierbarkeit nicht ausreichend präzise bestimmt. Das englische „control“ wird hier als Kontrolle übersetzt, aber diese Übersetzung ist ein wenig ungenau, da die im Deutschen bestehende Leitdifferenz zwischen Steuern und Regeln im Englischen weniger zentral ist – beide Vorgänge werden dort mitunter als „to control“ bezeichnet. Innerhalb der amerikanischen Terminologie ist der Verzicht auf diese Unterscheidung unkritisch, aber der hier angestrebte Vergleich mit dem Konzept der transklassischen Technik erfordert eine Differenzierung. Die später eingesetzten Begriffe des Beherrschens und Beschränkens fallen etwa formal unter das Regeln, während das Steuern unabhängig davon besteht¹²⁴. Gleichzeitig erfordert eine interdisziplinäre Darstellung des maschinellen Lernens keine detaillierte handlungstheoretische Aufarbeitung dessen, was unter Kontrolle verstanden werden kann und sollte. Insgesamt ergeben sich die folgenden Fragestellungen.

- Impliziert „Kontrolle“ eine aktive, nutzerseitige Entscheidung zur Übergabe eines speziellen Kontrollimpulses, wie etwa einer Lenkbewegung in einem Fahrzeug?

Abhängig davon, ob Formen un- oder vorbewusster Steuerung mitgedacht oder ausgeschlossen werden, ist der Raum der Technik, die kontrollierbar aber nicht wahrnehmbar ist, gegebenenfalls leer.

- Liegt schon „Kontrolle“ vor, wenn eine Überwachung des technischen Artefaktes möglich ist?

¹²⁴ Ganz abgesehen von Konzepten wie der Regelung durch Störgrößenaufschaltung.

Eine solche Überwachung setzt zumindest ein rudimentäres Verständnis der überwachten Technik voraus. Ein Beispiel zu dieser Form von „Kontrolle“ ist die Überprüfung des Ölstandes in einem Fahrzeug.

- Liegt bereits „Kontrolle“ vor, wenn der Nutzer des Artefaktes jederzeit einen Ausschalter, im Sinne eines Kill-Switches betätigen kann, um das Artefakt zu stoppen?

Computerviren erfüllen etwa das Kriterium der Unkontrollierbarkeit und das der Nicht-Wahrnehmbarkeit, können aber in den meisten Fällen durch physische Einwirkung auf die Stromversorgung oder die Netzwerkverbindung des befallenen Artefaktes an der Ausbreitung gehindert werden. Das bloße Ausschalten eines infizierten Systems würde intuitiv jedoch eher nicht als das Ausüben einer Kontrolle über die Computerviren angesehen werden.

Insgesamt scheint „Kontrolle“ als Kriterium etwas unscharf und als nicht unabhängig von potenzieller Unbeschränktheit. Ein alternatives Kriterium zur Kontrolle bildet das Verstehen. Das Verstehen wird zwischenzeitlich auch von Nordmann selbst explizit als Kriterium angedacht, er thematisiert es bereits im Abstract seines Textes als „human comprehension“.

[...] we no longer appear to perceive, comprehend, or control them, such that we no longer think of them as mechanisms or something “devised by human Wit”.

(Nordmann 2008, S. 176)

Die Einarbeitung dieser drei Fragestellungen in die Darstellung der naturalisierten Technik als Mengendiagramm resultiert unter Verwendung der Terminologie Nordmanns in der nachfolgenden Visualisierung.

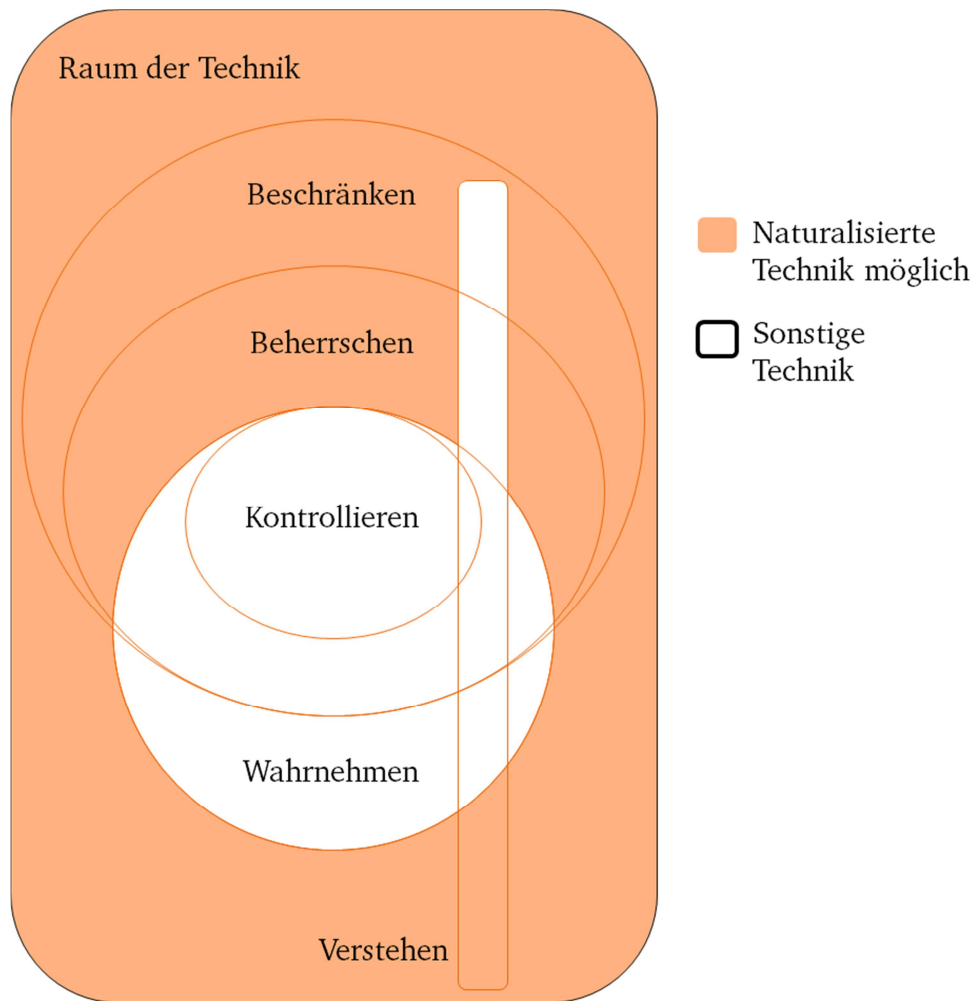


Abbildung 42: Raum naturalisierter Technik – Visualisierung B

Beschränken meint hier: Die Grenzen der Auswirkungen der Technik werden durch den Nutzer festgelegt. Ein Beispiel liegt darin, einem Kraftfahrzeug gezielt nur eine geringe Menge Kraftstoff zu überlassen und somit das Auftreten einer endlosen Kausalkette zu verhindern. Es ist in solch einem Fall nicht gefordert, dass irgendeine weitere Einwirkung auf das technische Artefakt möglich ist.

Beherrschen meint hier: Die Ziele beziehungsweise Zielwerte des technischen Artefaktes werden durch den Nutzer festgelegt. Gemeint ist keine unmittelbare, sondern eine mittelbare Steuerung, das heißt, die vorgegebenen Zielwerte werden von dem technischen Artefakt auf unbestimmte und gegebenenfalls auf selbsttätige oder autonome Weise erreicht. Die Ausformung der Vorgehensweise ist nicht notwendigerweise von den Einwirkungen des Nutzers abhängig und mitunter werden die vorgegebenen Ziele von dem technischen Artefakt mittels opaker Methoden erreicht. Diese mittelbare Steuerung besteht nicht darin, das Artefakt von Störgrößen abzuschirmen oder die unmittelbare Steuerung sicherzustellen. Stattdessen wird das Ziel des Artefaktes gesetzt und manipuliert. Im Beispiel des Kraftfahrzeuges wäre hier nicht nur die Menge an Kraftstoff, sondern auch das Fahrtziel durch den Nutzer vorgegeben.

Kontrollieren meint hier: Das Agieren des technischen Artefaktes bildet das antizipierte und gewollte Ergebnis der Einwirkungen durch den Nutzer. Am Beispiel des Kraftfahrzeuges bildet diese Stufe die tatsächliche Situation ab, in der die Menge an Kraftstoff und das Fahrtziel

durch den Nutzer fixiert sind. Der Nutzer hält darüber hinaus die Art der Erreichung dieses Ziel in den eigenen Händen und realisiert sie mittels klar überblickbarer Steuerungsimpulse. Ein anderes Beispiel wäre die Einleitung eines Bremsvorgangs, die ebenfalls einen Steuerungsimpuls darstellt und die Beherrschung und Beschränkung des Fahrzeugs impliziert¹²⁵. Die Sicherstellung von Steuerung findet in der graphischen Darstellung keine Entsprechung. Dies entspricht einer für den Kontext des maschinellen Lernens präzisierten Ausarbeitung der Begriffsbildung von „control“ bei Nordmann.

[...] we cannot control, what we cannot perceive.

(Nordmann 2008, S. 177)

Wahrnehmen meint hier: Das wirksame System als Ganzes, alle technischen Aspekte des Systems und alle Schnittstellen des Systems mit der Umwelt können prinzipiell wahrgenommen werden.

Verstehen meint hier: Die Wirkzusammenhänge können potenziell vollständig verstanden werden. Bei der Frage, ob ein Artefakt als naturalisierte Technik einzustufen ist, handelt es sich beim Verstehen genau dann um ein Ausschlusskriterium, wenn eine faktisch nicht unbeschränkte Technik durch das Verstehen der Zusammenhänge auch subjektiv als nicht unbeschränkt erscheint. Das bedeutet, wenn die Unbedenklichkeit der Langzeitwirkung eines Medikamentes nachgewiesen ist und dieser Nachweis bekannt wird, kann das Medikament nicht mehr als naturalisierte Technik betrachtet werden. Die genannte Charakterisierung naturalisierter Technik „[...] for all we know, a technical agency unfolds [...]“ greift nicht mehr, wenn verstanden ist, dass solch eine Annahme kontrafaktisch wäre.

Die hier skizzierten Begriffsoptionen sollen als detaillierte Kriterien zur Bestimmung naturalisierter Technik dienen, weisen allerdings noch immer Unklarheiten auf. Eine der Schwierigkeiten besteht in der Frage, ob im Falle einer mittelbaren oder medialen Steuerung überhaupt von Steuerung gesprochen werden sollte. Dies betrifft insbesondere die Manipulation von Selbstorganisationsprozessen im Nanobereich, aber auch im maschinellen Lernen gibt es Beispiele für mittelbare oder mediale Steuerung. Im instanzbasierten Lernen etwa kann die Reihenfolge der Eingabedaten gezielt manipuliert werden, während die Vorstruktur des MLA unbeeinflusst bleibt. In diesen Fällen kann weder wirklich von Steuerung noch von Regelung gesprochen werden. Die entsprechenden Bereiche nicht vorstrukturierter Technik scheinen sich sowohl aus Sicht der naturalisierten Technik als auch im Folgenden aus Sicht der transklassischen Technik einer Beschreibung zunächst zu entziehen. Die Wahl und Eigenschaften der Begriffsoptionen sollen jedoch nicht zu detailliert analysiert werden, da sie nur einen Zwischenschritt darstellen und die entstandene Präzisierung der ursprünglichen Visualisierung sich wieder vereinfachen lässt. Ein Beispiel für ein Potenzial zur Vereinfachung der Darstellung ist die explizite Betrachtung des Beherrschens. Der Grund für die Aufnahme des Beherrschens in das Mengendiagramm war, dass in der Rede von naturalisierter Technik diese Eigenschaft tendenziell in die Nähe von „kontrollierbar“ gesetzt wird. Die resultierende explizite Differenzierung zwischen Beherrschen und Beschränken bietet für das Mengendiagramm jedoch keinen Zugewinn an Aussagekraft. Die Visualisierung B lässt

¹²⁵ Unabhängig davon, dass der Steuerungsimpuls bei Glatteis auch ergebnislos bleiben kann.

demensprechend auch weniger komplex darstellen. Vor der Vereinfachung der Darstellung wird das Mengendiagramm zunächst etwas anders angeordnet. Die Visualisierung C stellt die neue Anordnung der Mengen dar, die sich nur optisch von der bisherigen Anordnung unterscheidet.

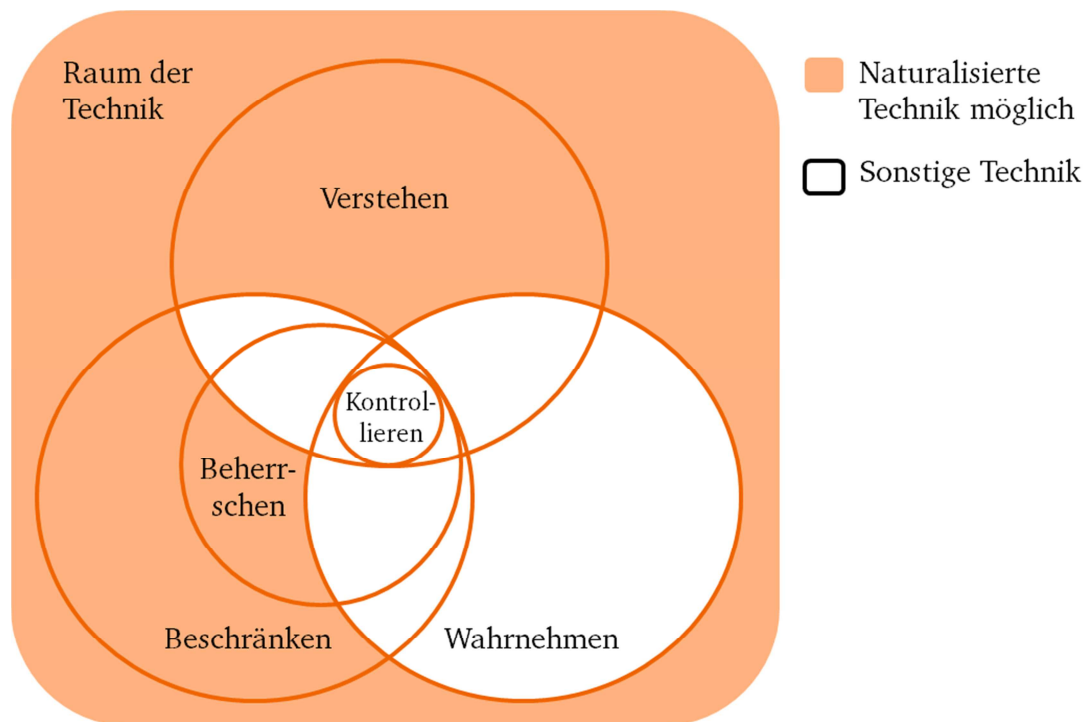


Abbildung 43: Raum naturalisierter Technik – Visualisierung C

Dieses Mengendiagramm kann nun bereinigt und vereinfacht werden, wodurch die nachfolgende Visualisierung entsteht.

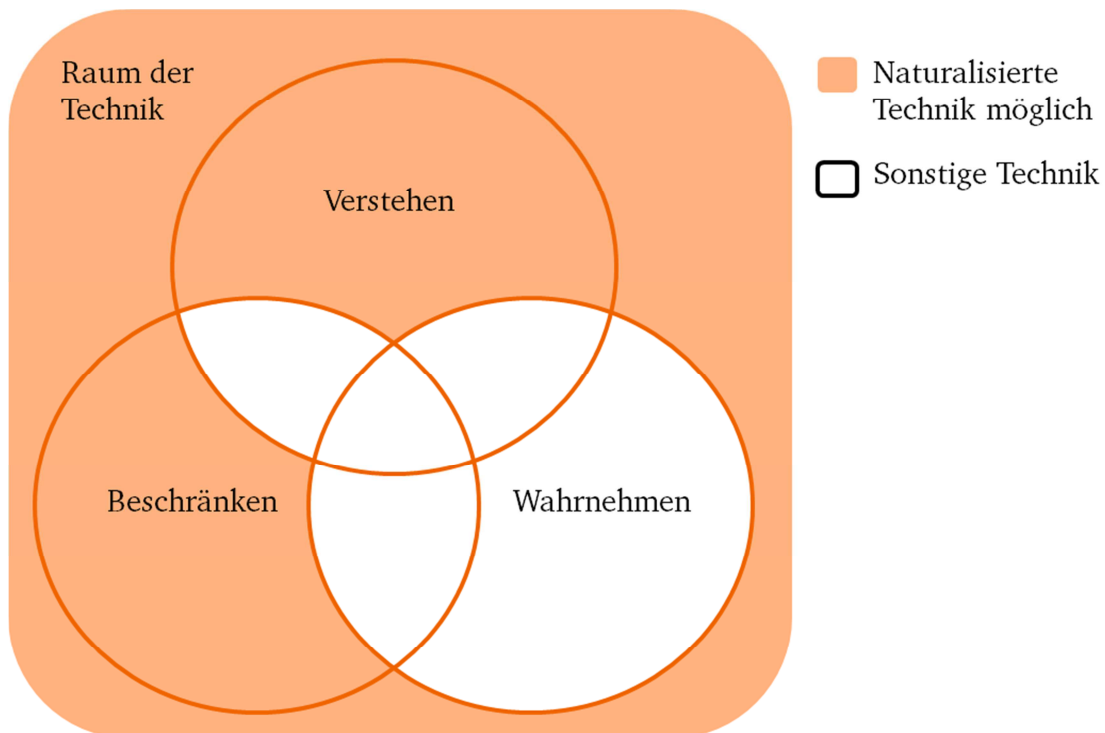


Abbildung 44: Raum naturalisierter Technik – Visualisierung D

Diese Darstellung zeigt zum einen, was aus Perspektive der naturalisierten Technik unter Kontrolle verstanden werden kann und inkludiert zum anderen die bei Nordmann selbst thematisierte, potenzielle Beschränktheit in die Darstellung. Weiterhin wird sich die naturalisierte Technik in dieser Darstellung mit der transklassischen Technik vergleichen lassen.

Anknüpfungspunkte maschinellen Lernens an naturalisierte Technik

Der Einsatz von maschinellem Lernen führt – implizit gewollt – dazu, dass Technik in verschiedenen Hinsichten unberechenbarer und undurchsichtiger wird. Diese Tendenz „entfernt“ die Technik je nach Abduktionsgrad (Kaminski et Harrach 2010) der eingesetzten Lernstrategien mit unterschiedlichem Nachdruck vom Bereich der Technik, die unmittelbar oder mittelbar gesteuert werden kann. Die Integration von maschinellem Lernen in eine Technik kann bezüglich der Kategorisierung als naturalisierte Technik einen Grenzübergang auslösen. Beispielsweise wird unmittelbar steuerbare Technik, das heißt wahrnehmbare und beschränkbare Artefakte, durch die Integration von selbsttätigen Autoadaptionsprozessen tendenziell zu lediglich beschränkbarer Technik. Gleichzeitig kann dieser Tendenz entgegengewirkt werden, indem etwa der theoretische Beweis der Konvergenz einer Lernstrategie erbracht wird. Diese Maßnahmen erhöhen das prinzipielle Verständnis der Einschränkungen, denen die Technik unterliegt und der Technik wird somit das Potenzial der Unbeschränktheit genommen. Die Technik wird sozusagen in den gemeinsamen Bereich von Beschränken und Verstehen zurückgedrängt.

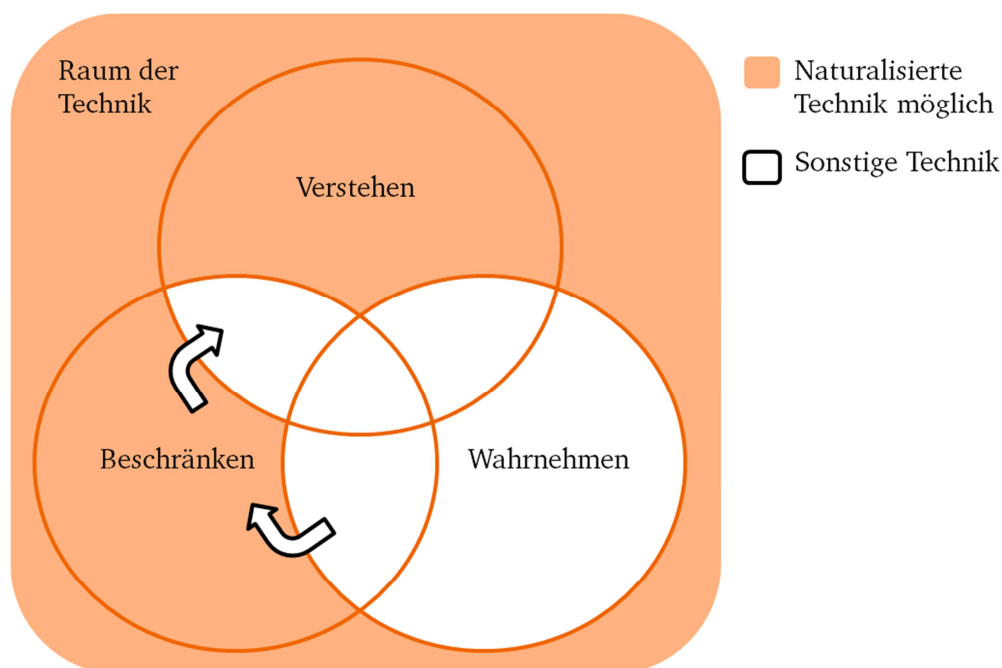


Abbildung 45: Grenzübergänge im Raum naturalisierte Technik

Maschinelles Lernen zu betrachten kann dementsprechend in der Diskussion und Bestimmung von naturalisierter Technik einen Mehrwert erbringen. Ein Nutzen der Verwendung dieser Perspektive für die Diskussion von maschinellem Lernen wiederum liegt in der Betrachtung der **Gegebenheit**¹²⁶ von Autoadaptionsprozessen.

¹²⁶ Die Frage nach dem Zweck oder Nutzen der Durchführung eines Autoadaptionsprozesses wird dabei bewusst zurückgestellt.

But when we black-box the working of a genetic modification or of automatic climate-control in a building, what remains is nothing at all but the technically altered environment itself that is indistinguishable in its mere givenness to a natural environment. Indeed this might serve as formal criterion for what are here called naturalized technologies: when you black-box it, there is nothing left.

(Nordmann 2008, S. 178)

Die Diskussion dieser Form von Gegebenheit wird im Folgenden nicht im Detail ausgearbeitet, sondern soll als offene Frage und möglicher Anknüpfungspunkt festgehalten werden. Die Kategorie der subsymbolischen Ansätze maschinellen Lernens bietet einen guten Startpunkt für solch eine Diskussion. Gerade subsymbolische Autoadaptionsprozesse im maschinellen Lernen bieten sich für eine Betrachtung besonders an, da sie automatisch eine Black Box darstellen, innerhalb derer unstrukturierte Rohdaten technisch verarbeitet werden. Der aus einem Autoadaptionsprozess resultierende Strukturvorschlag ist aus Sicht des Nutzers eine Art von veränderter Umgebung, denn auch wenn es vorher keine wahrnehmbare Struktur in der Umgebung gab, konnten doch die Rohdaten an das maschinell lernende Artefakt übermittelt werden. Diese Umgebung wird bewusst nicht als „technisch“ veränderte Umgebung bezeichnet, denn wenn etwa ein MLA das Schachspielen erlernt, dann hat sich dadurch zunächst nicht die Wahrnehmung von Schach durch den Nutzer des MLA verändert. Die Beobachtung der gespielten Zugfolgen kann jedoch analysiert werden und diese Analyse kann zu neuen Einsichten führen, die wiederum sehr wohl die Wahrnehmung des Nutzers von Schach verändert. Ein Beispiel ist die Vorbereitung von Eröffnungen¹²⁷.

It is hard to predict which opening will become fashionable, as computers now intervene into [opening] preparation and make it possible to prepare any opening in a limited amount of time.

(Rádío Xadrez im Interview mit Gelfand am 31.10.2012)

Diese Änderung wird aber eben nicht direkt technisch verursacht oder auch nur induziert, sondern der Nutzer hat einen Strukturvorschlag – beziehungsweise den durch ihn induzierten Weltbezug – interpretiert und analysiert, um die Umgebung selbst zu ändern. Insgesamt handelt es sich entsprechend nicht um eine technisch veränderte Umwelt. Die Methoden, die zur Umsetzung einer maschinellen Lernstrategie eingesetzt werden, sind hingegen durchaus als technisch beschreibbar. Die entsprechenden Autoadaptionsprozesse können ausgewählt, modifiziert und mit Methoden der Informatik effizienter gestaltet werden. Gleichfalls wird der Autoadaptionsprozess in einem technischen Artefakt implementiert und der Vorgang ist reproduzierbar. Diese vom Zweck des entstehenden Strukturvorschlages losgelöste Argumentation bezüglich des zugehörigen Autoadaptionsprozesses könnte auch bezüglich des

¹²⁷ Allerdings gibt es speziell für Schach durchaus auch die Gegenansicht, dass sich das generelle Verständnis seit Kasparov keineswegs weiterentwickelt hat und nur genau die Eröffnungen von der höheren Leistungsfähigkeit der Computer profitieren. „There has been practically no general development [after the Kasparov era]. The computer became stronger and so the analytical devices work better, so the opening analysis became more massive and more deep and so on, but it's not a development of some general understanding, or something like this, nothing.“ (Doggers im Interview mit Dvoretzky am 11.11.2010)

Aktes der Gestaltung eines **Kunstwerkes** geführt werden, der etwa anstelle der Bestimmung einer **Mutationsschrittweite** die Herstellung einer Farbe beinhalten kann. Der technische Charakter der Mittel, die zur Entstehung des Strukturvorschlages eingesetzt werden, überträgt sich entsprechend nicht ohne Weiteres auf den Strukturvorschlag selbst. Der Strukturvorschlag unterscheidet sich nicht von den Ergebnissen einer ebenfalls möglichen eigenen – das heißt menschlichen – Struktursuche des Nutzers. In beiden Fällen muss anschließend eine Konzeptualisierung beziehungsweise Interpretation der identifizierten Strukturen vorgenommen werden. Das im Strukturvorschlag Gegebene unterscheidet sich nicht davon, wie Nutzer sich selbst auf Basis von Rohdaten ihre Umgebung geben. Wenn Schachspieler auch ohne die Beobachtung eines Schachprogramms zu den entsprechenden Einsichten kommen können, ist entsprechend im Nachhinein nicht mehr erkennbar, ob die Einsicht mit Hilfe eines MLA oder eines Buches gewonnen wurde.

My success mainly has to do with the fact that I had the opportunity to learn more, more quickly. It has become easier to get hold of information. [...] Nowadays anyone can buy [recorded games] on DVD for 150 euros; one disk holds 4.5 million games. There are also more books than there used to be. And then of course I started working with a computer earlier than Vladimir Kramnik or Viswanathan Anand.

(Grossekathöfer im Interview mit Carlsen am 15.03.2010)

Die These an dieser Stelle ist, dass die gewonnene Einsicht keinen technischen Charakter mehr hat, sie ist eine Idee, wie Schach gespielt werden kann oder sollte. In Hinblick auf das Ergebnis unterscheidet sich damit die Struktursuche mit technischer Unterstützung nicht von einer rein menschlichen Suche, worin sich ein Beleg für die These findet, dass MLA Menschen in der Erstellung von deren Weltbezügen unterstützen können. Dies entspricht insofern nicht genau dem Betrachtungsfokus bei Nordmann, als der Strukturvorschlag im maschinellen Lernen meist durchaus als solcher erkennbar ist, auch wenn der Autoadaptionsprozess eine Black Box darstellt. Zusammengefasst wurde die These gestärkt, dass neugierige maschinell lernende Artefakte den Nutzer in der Erstellung eines Weltbezuges unterstützen und dass die erstellten Strukturvorschläge in dem Moment, in dem der Nutzer mit ihrer Unterstützung die Rohdaten wahrnimmt, nicht mehr von einem Strukturvorschlag rein menschlichen Ursprungs unterscheidbar sind.

Ein zweiter Gedanke, der aus der Diskussion naturalisierter Technik gewonnen werden kann, bezieht sich auf den Problembegriff der Selbstorganisation.

As we will see, [the limit where technology blends into nature and seemingly becomes one with it] could also be reached where engineering seeks to exploit surprising properties that arise from natural processes of self-organization.

(Nordmann 2008, S. 175)

Maschinelles Lernen passt auf den ersten Blick sehr gut zu dieser Argumentation der naturalisierten Technik, da die Nutzung unterschiedlich ausgeprägter Selbstorganisationsprozesse maschinelles Lernen gerade auszeichnet. Nicht-lernende Algorithmen arbeiten aus mathematischer Sicht auch bei der Verwendung von komplexen Selbstbezügen vergleichbar

zu einem Webstuhl oder Zahnradgetriebe – erst die Hinzunahme von Sensordaten und der Umgang mit diesen reduziert diese Vergleichbarkeit und ermöglicht die Parallele zur Selbstorganisation. Die Einbeziehung der Natur entlehnter Selbstorganisationsprozesse in die Diskussion beziehungsweise die initiale Entwicklung maschineller Lernstrategien wie evolutionärem Lernen oder künstlichen neuronalen Netzen wurde bereits dargestellt. Als wesentlich an der Einordnung durch Nordmann ist festzuhalten, dass das Nachbilden von Selbstorganisationsprozessen kein Alleinstellungsmerkmal von maschinellem Lernen ist, sondern dass auch andere moderne Technik sich ähnlich zu verhalten scheint. Dies gibt der Diskussion von maschinellen Lernstrategien einen größeren Rahmen und motiviert zusätzlich die Beschäftigung mit den Fragen, wie genau die Selbstorganisationsprozesse bei maschinellem Lernen kategorisiert werden können und was verschiedene maschinell lernende Artefakte genau leisten, beziehungsweise welcher Art von Technik sie zuzurechnen sind. Die technikahe Analyse der Nutzung von Selbstorganisationsprozessen innerhalb der Informatik kann die Diskussion anderer Technikbereiche vorbereiten, bei der bisher noch keine konkrete Technik systematisch diskutiert und entsprechend noch keine Grundlagenarbeit durchgeführt wurde.

Grenzen der Nutzbarkeit

Jenseits der genannten Möglichkeiten stößt die Perspektive der naturalisierten Technik bei der Diskussion von maschinellem Lernen schnell an Grenzen. Das Zusammentreffen von überraschenden Eigenschaften und Selbstorganisation findet im Rahmen des maschinellen Lernens gerade nicht statt. Diejenigen MLA, deren Strukturvorschläge erwartbare und damit potenziell überraschende Eigenschaften aufweisen, basieren wenig bis gar nicht auf Selbstorganisationsprozessen. Das Auftreten überraschender Eigenschaften weist stattdessen meist auf den zielorientierten Einsatz von MLA zur Lösung von Optimierungsproblemen hin. Gleichzeitig erwarten die Nutzer einer neugierigen Lernstrategie nicht, dass die Selbstorganisationsprozesse einen unmittelbaren Nutzen haben und können daher keine überraschenden Ergebnisse erhalten. Die Nutzer hoffen stattdessen, dass sie sich in ein nutzbringendes Verhältnis zu den Strukturvorschlägen setzen können. Die Diskussion maschinell lernender Algorithmen lässt sich entsprechend mit der Diskussion von Selbstorganisationsprozessen in anderen Gebieten verbinden, aber nicht darauf reduzieren. Eine zweite Grenze der Vergleichbarkeit liegt darin, dass die Diskussion des Einsatzes von maschinell lernenden Artefakten keine Diskussion möglicher Naturalismen erfordert. Die Frage, ob „Lern“-Prozesse im Rahmen von maschinellem Lernen natürlichen Lernprozessen oder gar menschlichen Lernprozessen entsprechen, steht nicht im Fokus – genauso wie die Vergleichbarkeit von KNN und menschlichen Gehirnen nicht betrachtet wird. Stattdessen wird darauf hingearbeitet, entscheiden zu können, wie stark die abduktiven Fähigkeiten eines MLA ausgeprägt sind und wie der Nutzer sich zu dem Artefakt in ein Verhältnis setzt. Eine weitere Unverträglichkeit maschinellen Lernens mit der Perspektive der naturalisierten Technik besteht darin, dass die Rede von naturalisierter Technik die Wahrnehmung eines Nutzers bezüglich eines genutzten Systems oder Artefaktes beschreibt. Der Eindruck, naturalisierter Technik in Form von maschinell lernenden Artefakten gegenüber zu stehen, kann nur in Ausnahmefällen entstehen, da genutzte Artefakte fast immer nur implizit auf maschinelles Lernen zurückgreifen und vom Nutzer nicht direkt über eine eigene Schnittstelle adressiert werden. Maschinelles Lernen ist nicht so sehr in sich bereits eine Technik als vielmehr ein Konstruktionsprinzip oder eine Hilfsttechnik, insbesondere da es implementiert werden muss und Charakteristika der Hardware, wie deren Nutzeroberfläche, großen Einfluss auf die

Wahrnehmung des resultierenden Artefaktes haben. Maschinelles Lernen kann allenfalls „naturalisierend“ wirken, wenn der Einsatz von Autoadaptionsprozessen die Möglichkeiten der Hardware zur Emulation von Natur erweitert. Darüber hinaus schließlich wird häufig nur der Strukturvorschlag analysiert und der Autoadaptionsprozess ist lange abgeschlossen oder zumindest eingefroren, wenn ein MLA betrachtet wird. MLA erfüllen in diesem Fall in Hinsicht auf die Autoadaptation nicht den zentralen Anspruch der potenziellen Unbeschränktheit. Die Grenzen der Nutzbarkeit der Perspektive der naturalisierten Technik bestehen insgesamt aus zwei wesentlichen Punkten. Zum ersten können MLA sich abstrakte Konzepte häufig gerade nicht über Lernstrategien aneignen, die besonders stark auf Selbstorganisationsprozesse und Neugier setzen. Das belegt noch einmal, dass maschinelles Lernen sich nicht auf die Umsetzung von Selbstorganisationsprinzipien reduzieren lässt. Der zweite wesentliche Punkt ist, dass bei der Analyse von MLA die Frage nach der gegebenen oder fehlenden Wahrnehmbarkeit nicht zentral ist, weshalb das maschinelle Lernen nicht im Fokus der naturalisierten Technik liegen kann. Das Fazit wird von der folgenden Textstelle gut zusammengefasst.

Technology naturalized is regressive in that it returns us to a state of ignorance towards our technical interventions that confront, perhaps dwarf us like uncomprehended nature.
(Nordmann 2008, S. 183)

Zwar liegt der zentrale Punkt maschinellen Lernens genau in der *Konfrontation* des Nutzers mit unvorhergesehenen und bis dahin noch unverstandenen Strukturen, andererseits erfolgt diese Konfrontation gezielt und transparent, um einen konstruktiven Umgang mit **Nichtwissen** bezüglich unzugänglicher Rohdaten zu ermöglichen. Der Einsatz von maschinellem Lernen ist eher eine Reaktion auf eine bereits identifizierte Ignoranz als ein Rückfall in eine solche.

3.5.2. MLA als transklassische Technik

Eine weitere Möglichkeit, moderne Hochtechnologie und damit maschinelles Lernen zu verorten, bietet sich mit der **transklassischen Technik** (Hubig 2008, S. 165-175; Hubig 2012). Hubig entwickelt seine Überlegungen zu transklassischer Technik im Kontext einer differenzierten Technikphilosophie, die im Zusammenhang mit einer Theorie von Technik als Medium, Mittel und Reflexionsbegriff steht und auf die im Kontext dieser Darstellung maschinellen Lernens nur hingewiesen sei (Hubig 2006; Hubig 2007).

Hubig fasst klassische Technik als das Gebilde, in dem die Grundfunktionen des **Steuerns** und **Regelns** wahrgenommen werden. Diejenige des Regelns dadurch, dass die natürliche **Medialität** künstlich überformt wird, da eine Steuerung im Rahmen natürlicher Medien immer der Kontingenz von Störgrößen aus der Umwelt dieser Medialität unterliegt – beispielsweise dem Wetter oder den Wanderbewegungen des Wildes. Klassische Technik besteht bei Hubig entsprechend darin, Regelungsmechanismen, das heißt Systembildungen, mittels einer technischen Überformung der natürlichen Medialität zu realisieren¹²⁸. Anders ausgedrückt wird klassische Technik handlungstheoretisch konzipiert, Handlungen und die Ausbildung von Handlungskompetenz werden an der Differenz von vorgestelltem und

¹²⁸ Sowohl der äußeren als auch der inneren Medialität, das heißt der sozialen Natur der Menschen.

realisiertem Zweck orientiert und Technik sichert dabei die Erwartbarkeit von Handlungserfolgen ab. Diese notwendigen Bedingungen erfüllt die transklassische Technik nicht mehr, hier fällt die Art der technischen Überformung der Medialität anders aus. Insbesondere sind die Prozesse nicht mehr disponibel, das heißt, nicht mehr durch die Subjekte gestaltbar. Die Nutzer haben keine Repräsentationen der Technik und können keine solchen mehr bilden, sie können nicht intervenieren, sei es steuernd oder regelnd – die Schnittstellen verschwinden. Diese Definition transklassischer Technik betrifft Entwickler und Nutzer als die Subjekte, die mit der Technik umgehen. Insgesamt deutet die Vorsilbe „trans“ in einem schwachen Sinne ein „jenseits“ der klassischen Technik an. Diese Definition ist deutlich zu unterscheiden von einer positiven Definition transklassischer Technik wie sie Bense (Bense et Walther 1998) gibt. Bense betrachtet etwa die Kerntechnik als eine transklassische Technik, die nach der Definition Hubigs eine klassische Technik darstellt. Der Begriff transklassischer Technik entspricht ebenso nicht dem Begriff bei Günther oder Wiener.

Visualisierung naturalisierter Technik

Die Visualisierung der naturalisierten Technik wurde im Vorherigen bewusst so gewählt, dass die nachfolgende Diskussion transklassischer Technik auf dem gleichen Mengendiagramm aufbauen kann.

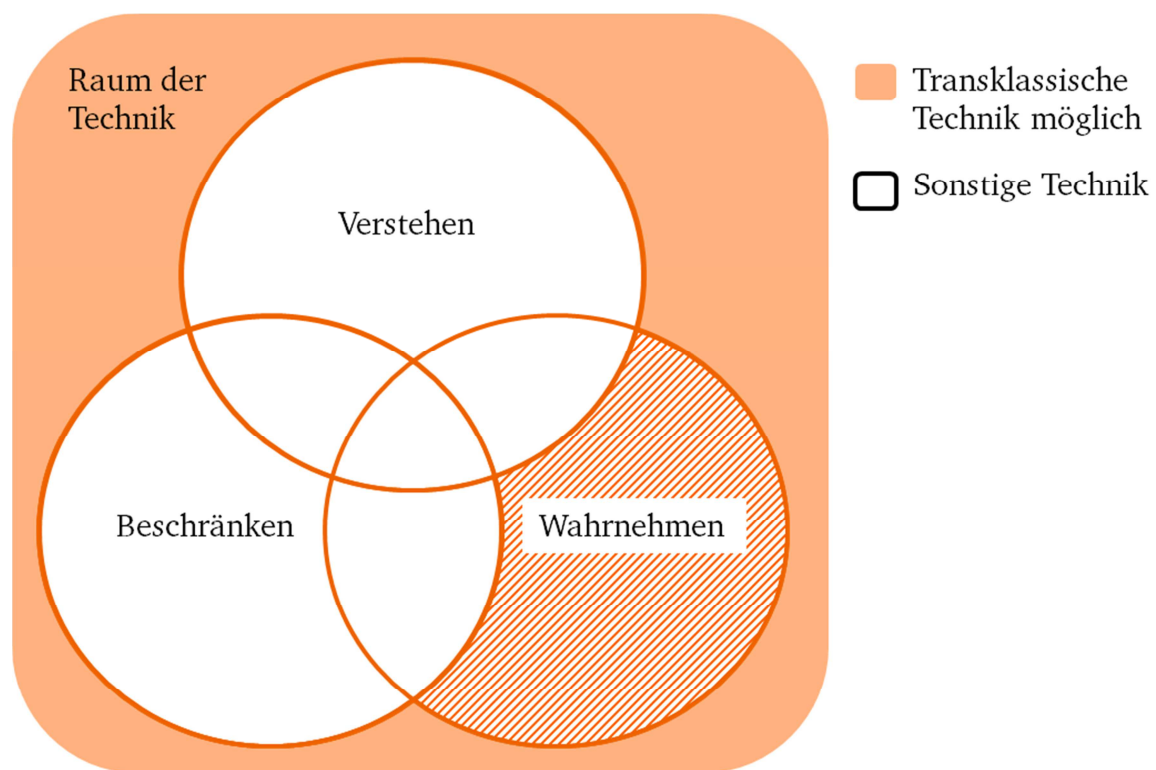


Abbildung 46: Raum transklassischer Technik

Beschränken meint hier: Die Grenzen der Auswirkung der Technik werden durch den Nutzer festgelegt. Um die Kriterien transklassischer Technik zu erfüllen, muss eine Technik in der Lage sein, aktiv die Handlungsoptionen des Nutzers einzuschränken oder festzulegen.

Verstehen meint hier: Die Konsequenzen der Einwirkungen des Nutzers sind transparent.

Wahrnehmen meint hier: Alle Schnittstellen des Systems mit der Umwelt *und die eigentliche Umwelt* können vom Nutzer wahrgenommen werden. Das Wissen um die Vollständigkeit kann

aus dem Verstehen kommen oder daraus, dass zusätzlich das wirksame System als Ganzes wahrgenommen werden kann. Die Einbeziehung der Wahrnehmbarkeit der eigentlichen Umwelt lässt sich nicht ohne Weiteres in die Visualisierung der naturalisierten Technik übertragen, daher ist der entsprechende Bereich schraffiert dargestellt.

Anknüpfungspunkte maschinellen Lernens an transklassische Technik

Eine erste Parallele bezüglich der Diskussion maschinellen Lernens und der genannten Definition transklassischer Technik liegt in der Berücksichtigung der Wahrnehmbarkeit der Umwelt. Die Umwelt ist beim Einsatz eines MLA meist gerade nicht zufriedenstellend als Umwelt wahrnehmbar. Die Umwelt entspricht im Falle des Bestehens von Vorvermutungen bezüglich der Struktur der Umgebung einer Menge von Rohdaten, die formal die vermutete Struktur instanzieren, aber kein intelligibles Gesamtbild ergeben. Im Falle des Fehlens solcher Vorvermutungen bezüglich der Struktur setzt sich die Umwelt lediglich aus einer Menge kontingent erscheinender Rohdaten zusammen. Genau diese fehlende Wahrnehmbarkeit der Rohdaten motiviert häufig den Einsatz maschinell lernender Artefakte und liegt entsprechend bei maschinellem Lernen häufig auch unabhängig von der eingesetzten Technik vor. In Konsequenz kann eine Technik, die maschinelles Lernen verwendet, zu Beginn des Autoadaptionsprozesses vollständig verstanden sein und nach dem Durchlaufen einiger Iterationen des Autoadaptionsprozesses dennoch die Kriterien transklassischer Technik erfüllen. Dies kann etwa geschehen, wenn die zu erlernenden Rohdaten zunächst transparent, aber unstrukturiert vorliegen, im Rahmen der Autoadaptionsprozesse jedoch nicht voraussehbare beziehungsweise verstandene Veränderungen der Zusammenhänge innerhalb des lernenden Artefaktes entstehen.

Eine zweite Parallele bezüglich der Diskussion maschinellen Lernens und der genannten Definition transklassischer Technik besteht darin, dass der Übergang von einer klassischen Vorstellung von Technik

Handlungskompetenz als Fähigkeit der Zweckrationalisierung entwickelt sich [...] auch und gerade als externe Fähigkeit (vom Beobachterstandpunkt aus), zwischen [medialen] Voraussetzungen die adäquate auszuwählen oder die Voraussetzungen höherstufig selbst zu gestalten und weiterzuentwickeln.

(Hubig 2008, S. 167)

zu einer transklassischen Vorstellung

Diese Vorstellung „klassischer Technik“ [...] wird nun durch Entwicklungen „transklassischer Technik“ entscheidend relativiert und herausgefordert. Maßgeblich hierfür erscheinen diejenigen Hochtechnologien, die unsere innere und äußere Natur „technisieren“, „technisch überformen“, sowie mit „autonomer“ Problemlösekompetenz versehen:

[...]

Im Zuge der neuen Entwicklungen nun scheinen die Schnittstellen, wenngleich sie objektiv nicht verschwinden, so doch in gewisser Hinsicht indisponibel zu werden [...]

(Hubig 2008, S. 167)

sich zumindest für *zielorientiert* maschinell lernende Artefakte gut nachvollziehen lässt. Die Beobachtung der reduzierten Disponibilität von **Schnittstellen** lässt sich darüber hinaus auch auf *neugieriges* maschinelles Lernen übertragen. Die Schnittstellen zu neugierigen Artefakten werden insofern indisponibel, als die aus den Lernvorgängen resultierenden Strukturvorschläge nicht reversibel sind. Die Vorschläge sollen und können zunächst nur wahrgenommen werden, allerdings kann eine einmal erfolgte Wahrnehmung anschließend nicht zurückgenommen werden. Ein Blick auf die Rohdaten nach Betrachtung des Strukturvorschlages ist ein anderer als zuvor. Die prinzipielle Vorstellbarkeit von Mitteln¹²⁹ ist entsprechend in den neugierig maschinell lernenden Artefakten schon angelegt, die Ausformulierung der Mittel allerdings nicht, diese erfolgt erst im Rahmen der nutzerseitigen Interpretation der Strukturvorschläge. Auch wenn hier die Mensch-Technik-Schnittstelle zu einer Kopplung wird, werden den Nutzern keine fertigen „Bilder einer Welt [...] vermittelt“ (Hubig 2008, S. 168). Anders formuliert weisen neugierige MLA zwar *einerseits* eine Kopplung und keine echte Schnittstelle auf, weil die Strukturvorschläge nur präsentiert werden. Nutzer können die Strukturvorschläge wahrnehmen, sind dann aber gezwungen sie auch zu interpretieren, da sie Eindrücke nicht „nur“ wahrnehmen können. Die einzige Alternative liegt in diesem Fall darin, die Strukturvorschläge zu ignorieren, diese Wahl stellt jedoch mangels Bezug zum Autoadaptionsprozess des MLA keine Nutzung einer Schnittstelle dar. *Andererseits* erhalten die Nutzer beim Einsatz neugieriger und, mit Abstrichen, zielorientierter MLA gerade *keine* „fertig“ konzeptualisierten Ergebnisse. Die Diskussion des Umgangs mit den Schnittstellen und der Bedeutung der Transparenz von Technik – oder zumindest des Offenlegens eines systematischen Fehlers und des Suchraums eines Autoadaptionsprozesses – lässt sich dementsprechend auf beide Varianten maschinellen Lernens übertragen.

Anknüpfungspunkte speziell bei zielorientiertem maschinellern Lernen

Eine darüber hinausgehende Nutzbarmachung der Perspektive der transklassischen Technik basiert darauf, dass eine Diskussion von Weltbezügen eröffnet wird, die sich sehr gut auf zielorientiert maschinell lernende Artefakte übertragen lässt. Dies wird insbesondere in der Analyse von theoretischen und praktischen Weltbezügen deutlich.

¹²⁹ Details zur Vorstellbarkeit von Mitteln finden sich in „Der technisch aufgerüstete Mensch. Auswirkungen auf unser Menschenbild“ (Hubig 2008, S. 171).

[...] die Medialität des Technischen wird in einer Weise „selbstverständlich“, die nicht mehr erlaubt, jenseits ihrer konkurrierende Weltbezüge positiver oder negativer Art (als Differenzerfahrungen) wahrzunehmen und zu gestalten.

(Hubig 2008, S. 171)

Der Schwerpunkt dieser Formulierung entspricht nicht ganz demjenigen der hier vertretenen Perspektive auf maschinelles Lernen, aber dennoch kann so über Weltbezüge gesprochen und die Idee konkurrierender Weltbezüge formuliert werden. Als Differenzerfahrung konkurrierende Weltbezüge wahrzunehmen und zu gestalten, entspricht genau einem möglichen Wunsch bei der Erzeugung von Strukturvorschlägen. Zielorientierte MLA lassen sich weiter fast ohne Verluste unter der hier definierten zielorientierten Technik verorten.

Die ehemals funktionsorientierte Technik wird, so die Forderung, zu einer zielorientierten Technik, die auf einer adaptiv gewonnenen Informationsbasis antizipatorisch die Problemlösungen vornimmt [...].

(Hubig 2008, S. 172)

Die Fragestellungen im Umgang mit transklassischer Technik und Lösungsansätze wie derjenige der **Parallelkommunikation** lassen sich entsprechend sehr gut auf Implementierungen von zielorientierten Lernstrategien übertragen. Eine auf dem Konzept der Parallelkommunikation aufbauende Möglichkeit, transklassische **Technikkompetenz** im Einsatz von zielorientiertem maschinellem Lernen zu fördern, besteht darin, randomisierte Sensordaten an das MLA zu übergeben und die resultierende Autoadaption zu beobachten¹³⁰. Wenn ein Autoadaptionsprozess einige Iterationen durchlaufen und damit bereits eine **Vorstruktur** zur Bildung von Strukturvorschlägen entwickelt hat¹³¹, und dem Artefakt anschließend zufällige Eingabedaten übergeben werden, kann das zielorientierte MLA durch eine Adaption in Reaktion auf diese zufälligen Daten seinem Ziel nicht näher kommen. Die Zufälligkeit der Eingaben erlaubt es der Technik nicht, sich zu Gunsten ihrer Zielvorstellungen zu adaptieren. Der derzeitige Zustand wird dementsprechend in Hinsicht auf die Performanz annähernd erhalten und dennoch wird eine Veränderung des Artefaktes bewirkt. Diejenigen Aspekte beziehungsweise Teile der erlernten Vorstruktur, die sich innerhalb des Artefaktes durch die Übergabe zufälliger Sensordaten verändern, sind daher in Hinblick auf das formulierte Ziel wahrscheinlich nur von geringer Bedeutung. Durch solch eine Manipulation der Sensordaten können gegebenenfalls die stabilen Teile der resultierenden Vorstruktur gezielt als konzeptuell relevant identifiziert werden. Diese Vorgehensweise wäre ein möglicher Ansatz für einen Umgang mit Nichtwissen bezüglich der Vorstruktur und des Strukturvorschlages. Ein solches absichtsvolles Verrauschen der Eingabedaten kann auch durchgeführt werden, wenn das Artefakt Autoadaptationen auf Basis der Eingaben des Nutzers durchführt, das heißt, falls **Nutzerstereotypen** erstellt werden. Diese Variation des Ansatzes entspricht der dritten Ebene der Parallelkommunikation, in der ein Austausch verschiedener

¹³⁰ Details zur Diskussion von Technikkompetenz – insbesondere in Hinblick auf Hubigs Perspektive – finden sich bei Sesink (Sesink 2011, S. 11f).

¹³¹ Diese Formulierung wird in der Diskussion nichttrivialer Maschinen wieder aufgegriffen und genauer ausgearbeitet.

Nutzer der gleichen Technik angeregt wird, um den derzeitigen Zustand der Technik zu bestimmen. Im Kontext eines MLA ist solch ein Austausch besonders gut in der Praxis durchführbar, da das Artefakt sich an die unterschiedlichen Nutzer anpasst und so der Zugriff verschiedener Nutzer automatisch eine Form von mittlerem Nutzerstereotyp erstellt. Auf diese Weise kann nicht nur der Zustand und die Reaktionsweise des MLA besser überblickt werden, sondern es kann je nach Umfeld des Einsatzes auch dessen Überanpassung an die Eingaben des Nutzers reduziert werden – etwa wenn das MLA eine Aufgabe durch Beobachtung des Nutzers lernen soll und bei der Beobachtung einer Gruppe die individuellen Fehler verwirft.

Grenzen der Übertragbarkeit auf maschinelles Lernen

Größere Unstimmigkeiten bei der Betrachtung von maschinellern Lernen als transklassische Technik treten nur bei neugierigen MLA auf, da die zu Grunde liegenden Algorithmen zentrale Schwächen transklassischer Technik nicht oder zumindest in deutlich unterschiedlicher Form aufweisen. In erster Linie stellen die Strukturvorschläge keine Repräsentationen von Konzepten, sondern allenfalls eine Präsentation der eingelesenen Rohdaten dar und somit besteht das folgende zentrale Problem transklassischer Technik in dieser Form nicht.

Durch die im Zuge des Ubiquitous Computing vollzogene Intellektualisierung der Handlungsumgebungen [wird] der Zustand herbeigeführt, dass die Strategien der Identifizierung der Elemente der Handlungsumwelt [...] in den IT-Systemen selbst implementiert sind, mithin unsere Real- und Sozialtechnik letztlich durch eine in die Systeme verlegte Intellektualtechnik (als Umgang mit Repräsentationen) dominiert würde.

(Hubig 2008, S. 167)

Die Repräsentation der Strukturvorschläge erfolgt durch den Nutzer. Das heißt, die Technik dient in diesem Fall dazu, die Erzeugung oder Entdeckung von Welt zu ermöglichen. Alle Welterfahrung, die auf diese Weise gemacht werden kann, muss der Nutzer selbst initiieren.

Unsere Welterfahrung baut sich auf der Wahrnehmung der Differenz zwischen dem vorgestellten (prognostizierten) und dem realisierten Zweck auf [...].

(Hubig 2008, S. 166)

Es handelt sich bei neugierig lernenden Artefakten nicht notwendig um konzeptuell reiche Repräsentationstechnik wie sie etwa bei einem Retinaimplantat vorliegt, das Daten in einer sehr spezifischen Struktur aufbereiten muss, so dass die Daten als optische Signale interpretiert werden können. Neugierige MLA entsprechen in dieser Hinsicht gerade nicht solchen bildgebenden und ähnlichen Verfahren, die bei Hubig unter dem Begriff der **Simulation** zusammengefasst werden.

[...] im Bereich der Simulationen und bildgebenden Verfahren, die je nach verarbeiteter Datenmenge und -qualität, berücksichtigten Parametern und unterstellten Kausalmodellen uns Sachlagen präsentieren [...].

(Hubig 2008, S. 169)

Neugierig maschinell lernende Artefakte sind, wie bereits diskutiert wurde, keine **Informationstechnik** im eigentlichen Sinn. Sie verarbeiten nicht Informationen, sondern Rohdaten ohne Struktur, denen nur das Potenzial zugeschrieben wird, auch als Trainingsdaten nutzbar zu sein. Die entstehenden Strukturvorschläge sollen mitunter den Nutzer dabei unterstützen Weltbezüge zu entwickeln beziehungsweise Welt zu entdecken, vor allem bei Einsatz von neugierigen Autoadaptionsprozessen. Diese Form von MLA wird daher in Abschnitt 3.7.2 als **Welttechnik** anstatt als Informationstechnik bezeichnet werden. Eine solche Form von Technik weist keine Problemlösekompetenz auf. Sie passt sich zwar an Rohdaten an, reagiert dabei aber nicht auf Probleme, sondern auf Basis ihrer Vorstruktur auf Eingaben, die nicht als Informationen, sondern als Reize bezeichnet werden sollten.

Die schlechte transklassische Verortbarkeit von speziell neugierigem maschinellern Lernen lässt sich am besten in einem Vergleich mit den Kriterien zur Einordnung einer Technik als **Biofakt** (Karafyllis 2003) diskutieren.

Biofakte beruhen darauf, dass Wachstum und Reproduktion technisch induziert sind.

[...]

[Wir finden] die Implementation von „natürlichen“ Strategien, unabhängig von Stoffen und Gesetzen in den Simulationen von Entwicklungs- und Reproduktionsprozessen, wie sie die Wachstums- und Evolutionsforschung vornimmt. Gemeinsam ist diesen (hier nur grob unterschiedenen) Implementationsformen, dass – wenn auch im Ergebnis nicht mehr disponibel oder revidierbar – die technische Induzierung rekonstruierbar bleibt.

Die eigentlichen Biofakte beruhen hingegen auf einer Fusion von Technik und Natur. Eine echte Fusion liegt vor, wenn Wachstums- und Reproduktionsprozesse technisch provoziert oder stimuliert werden, wobei im Ergebnis der technische oder natürliche Anteil nicht mehr zu sondern ist.

(Hubig 2008, S. 168)

Im Kontext des maschinellen Lernens wird gezielt nicht der Begriff des **Cyberfaktes** genutzt, da dieser Begriff stark in Hinsicht auf Technik aus dem Ubiquitous Computing – wie Nutzerstereotypen und Unsichtbarkeit – geprägt ist und systematisch erweitert werden müsste, um maschinellern Lernen mit abzubilden. Maschinellern Lernen ist bezüglich der Diskussion der Cyberfakte schlicht von einer geringen Relevanz, weswegen die Argumentation im Weiteren stattdessen mit Hilfe der Idee der Biofakte geführt wird, um irreführende Assoziationen zu vermeiden, die aus der begrifflichen Nähe der Rede von den Cyberfakten stammt. In Hinsicht auf Biofakte scheint ebenfalls bereits im Vorhinein klar zu sein, dass es sich bezüglich maschinellen Lernens um eine Sackgasse handelt, allerdings birgt die

Betrachtung der Biofakte, wie schon andere Teile dieser Diskussion, dennoch einen erkennbaren Gewinn. In diesem Fall liegt der Gewinn in erster Linie in der Erkenntnis, woran genau die Übertragung scheitert und beleuchtet damit, wo genau die Schwierigkeiten liegen, wenn neugieriges maschinelles Lernen als transklassische Technik verortet werden soll.

Die Einordnung einer Technik als Biofakt erfordert, wie oben beschrieben, eine Ununterscheidbarkeit des technischen und natürlichen Anteils. Die Strukturvorschläge neugierigen Lernens sind zunächst klar als künstlich und damit in diesem Kontext als technisch identifizierbar. Nachdem der Nutzer sich jedoch in ein Verhältnis zu den Strukturvorschlägen gesetzt hat, ist rückblickend nicht mehr unterscheidbar, ob ihm die Interpretation der Struktur auf einem natürlichen Weg oder technisch präsentiert oder vermittelt aufgefallen ist. Die resultierende Umgebung ist nicht technisch gestaltet, sie wird erst durch eine Interpretation des Nutzers möglich und ist veränderbar. Der Nutzer wünscht sich bewusst einen Strukturvorschlag, den er aus seinem – üblicherweise von den Rohdaten unabhängigen – Vorwissen heraus interpretieren kann. Das heißt, eine Interpretation kann nur stattfinden, wenn der Nutzer Zugang zu einer anderen Wahrnehmungsstruktur als der vorgeschlagenen hat und wenn diese Struktur in einen Bezug zu dem Strukturvorschlag gesetzt werden kann. Dieses In-Bezug-Setzen scheitert auch tatsächlich in der Praxis in der Mehrzahl der Fälle, da die Strukturvorschläge sehr häufig als irrelevant betrachtet werden oder aus anderen Gründen nicht genutzt werden können. Im Rahmen von neugierigem maschinellem Lernen wird entsprechend vom Nutzer durchaus das Fehlen eines „konkurrierenden“ Weltbezuges festgestellt und das maschinell erlernte Strukturangebot als konkurrenzlos und unabweisbar akzeptiert, aber das Angebot wird doch immer nur als ein unterscheidbares Hilfsangebot zur Erstellung eines Weltbezuges interpretiert. Wenn beispielsweise ein Autohaus im Februar des vergangenen Jahres sehr viele Kunden gewinnen konnte, deren Nachnamen mit einem „H“ beginnen, so wird dieser Strukturvorschlag vom Betreiber des Autohauses wahrscheinlich als zufällig abgetan. Das bedeutet, der Betreiber kann keine Interpretation für den Strukturvorschlag finden, die auf eine ausbeutbare Systematik hindeuten würde und entscheidet, dass er den Strukturvorschlag verwirft. Der Betreiber konnte in diesem Fall nicht auf eine andere Struktur rekurren, obwohl durchaus eine Vielzahl solcher Systematiken denkbar ist. Das im Februar mit Rabatten beworbene Modell könnte etwa ebenfalls mit „H“ begonnen haben oder eine besonders erfolgreiche Form der Werbung könnte nur einem Teil seiner Kundschaft zugestellt worden sein. Unabhängig davon, ob im Einzelfall eine interessante, systematische Ursache für die Entstehung des Strukturvorschlag identifiziert werden kann, besteht der Mehrwert des maschinellen Lernens genau darin, dass die Erstellung der Strukturvorschläge automatisch durchgeführt werden kann. Ein MLA kann die Strukturvorschläge – die unabhängig von ihrer Interpretierbarkeit zumindest zu einem Großteil auf faktisch vorliegenden Strukturen in den Rohdaten basieren¹³² – extrem schnell erstellen. Unverständliche Strukturvorschläge können entsprechend ohne Bedenken verworfen werden, da in kurzer Zeit eine Alternative generiert werden kann. Unabhängig von der Frage nach der Unterscheidung des technischen und natürlichen Anteils können die „Bedingungen des Wirkens (Hubig 2008, S. 168) neugierig maschinell lernender Artefakte aufgrund der meist opaken Strategien und der fehlenden Struktur in den Rohdaten häufig nicht beeinflusst werden, weil sie nicht verstanden werden. Die Bedingungen des Wirkens

¹³² Wie im ersten Hauptteil dargestellt, verfälschen manche Autoadaptionsprozesse ihre Eingaben, Vorstrukturen und Strukturvorschläge gezielt, um Überanpassungen vorzubeugen. Allerdings wird in diesen Fällen nur versucht, die Strukturvorschläge weniger spezifisch zu machen und nicht sie rundweg falsch reagieren zu lassen.

können häufig lediglich ungezielt manipuliert werden, indem etwa ein Teil der verfügbaren Rohdaten nicht eingespeist wird. Die Schwierigkeiten bei der Manipulation der Bedingungen des Wirkens übertragen sich nicht ohne Weiteres auf die Manipulation des eigentlichen Wirkens eines MLA. Je nachdem wie das Wirken gefasst wird, kann dieses sogar im Rahmen der anschließenden Interpretation der Strukturvorschläge beeinflusst werden. Zusammengefasst liegen die beiden Abweichungen zu Biofakten vor allem in der speziellen Situation bezüglich der Unterscheidung zwischen technischem und natürlichem Anteil, sowie in der Möglichkeit, das Wirken der MLA zu beeinflussen, während die Bedingungen des Wirkens sich der systematischen Manipulation tendenziell entziehen.

Abschließend kann festgehalten werden, dass die Entscheidung, ob zielorientiertes maschinelles Lernen als der transklassischen Technik zugehörig betrachtet werden sollte, an dieser Stelle noch nicht geklärt wurde. Eine Aussage ist nur bezüglich der neugierigen Autoadaptionsprozesse möglich, deren Integration die Schärfe des Begriffes der transklassischen Technik reduzieren würde. Die Herausforderungen bei der Betrachtung von neugierigen MLA entsprechen nicht der ursprünglichen Stoßrichtung der Rede von transklassischer Technik – der Aufdeckung und Bearbeitung der Problemfelder, die aus dem Verlust der Spuren resultieren. Herausforderungen dieser Art bestehen für neugieriges Lernen gerade nicht und entsprechend sollten neugierige Artefakte tendenziell nicht als transklassische Technik verortet werden. Es würde sich darüber hinaus anbieten, für andere Technik aus dem Umfeld der **NBIC**-Technik ebenfalls zu überprüfen, inwieweit sie die Problemlage der transklassischen Technik betrifft und welche Teilbereiche Ausnahmen bilden, die exkludiert werden sollten. Gegebenenfalls können sich aus den Gemeinsamkeiten dieser Ausnahmen Strategien gewinnen lassen, die einen besseren Umgang mit transklassischer Technik generell oder zumindest mit den Ausnahme- und Übergangsbereichen ermöglichen. Umgekehrt kann auch die Frage gestellt werden, inwiefern das neugierige Lernen sich aus der Perspektive der transklassischen Technik der klassischen Technik zuordnen lässt. Das neugierige maschinelle Lernen setzt das nutzerseitige Sich-in-ein-Verhältnis-Setzen zu den Ergebnissen des Prozessierens gerade nicht außer Kraft, sondern es ermöglicht es erst – ja erzwingt es sogar. Mit dieser Ermöglichung liegt ein begriffliches Merkmal klassischer Technik in neuer Form vor. Die ursprüngliche Frage dieses Abschnittes war, wie sich maschinelles Lernen aus der Perspektive der transklassischen Technik verorten lässt. Die zielorientierten Autoadaptionsprozesse ließen sich aus der Perspektive der transklassischen Technik sehr gut einordnen, während dies für den Bereich der neugierigen MLA weniger gut gelang. Festzuhalten ist darüber hinaus jedoch, dass sich neugieriges maschinelles Lernen an der transklassischen Technik gespiegelt nicht nur nicht als transklassisch verorten lässt, sondern sogar eine besondere Form *klassischer* Technik zu sein scheint.

3.5.3. MLA als nichttriviale Technik

Eine weitere interessante technikphilosophische Perspektive, die auf ihre Eignung zur Beschreibung maschinellen Lernens geprüft werden kann, ist die von Heinz von Foerster stammende Unterscheidung trivialer und **nichttrivialer Maschinen** (Foerster 2006, S. 6ff; Foerster 1993, S. 245ff). Von Kaminski (Kaminski 2012, S. 6ff) wurde bereits diskutiert, wie der Begriff der Nichttrivialität erweitert werden kann, um maschinelles Lernen besser abdecken zu können und diese Betrachtungen sollen hier als Grundlage dienen.

Eine Schwierigkeit für die Diskussion des maschinellen Lernens ergibt sich daraus, dass Kaminski gezielt Technikbereiche betrachtet, in denen mit den Artefakten informell

umgegangen wird (Schmidt 2007) oder in denen den Nutzern gänzlich unbekannt ist, dass maschinelles Lernen Verwendung findet (Isermann et al. 2009). Diese Anwendungen fallen in den Bereich des Ubiquitous Computing oder der **smarten** Artefakte¹³³ und stehen daher weniger im Zentrum einer Analyse des maschinellen Lernens. Insbesondere gilt, dass die Nutzer sich bei dieser Art von Artefakten häufig zu Strukturvorschlägen in ein Verhältnis setzen, bei denen der zugrunde liegende Autoadaptionsprozess bereits abgeschlossen ist. Wie bereits diskutiert wurde, können Strukturvorschläge durchaus interpretiert werden, ohne dass der zu Grunde liegende Autoadaptionsprozess noch aktiv ist und dies stellt in der Praxis durchaus den Normalfall dar. Ein den Nutzer unterstützender Autoadaptionsprozess – wie bei der von Kaminski als Beispiel genannten, sich stetig selbst verbessernden Spracherkennung – ist für neugieriges maschinelles Lernen eher die Ausnahme, weil die Nutzer in solch einem Szenario ständig die Anpassungen durch den Autoadaptionsprozess des Artefaktes bewerten müssen. Dies ist in vielen Anwendungsfällen des neugierigen Lernens nicht praktikabel, da Versuche der Interpretation von Strukturvorschlägen meist aufwendig sind und häufig fehlschlagen. Vielfach ist das Gelingen eines solchen Interpretationsversuches Anlass für die Beendigung des entsprechenden Autoadaptionsprozesses. Insgesamt zeigt diese Schwierigkeit jedoch nur einen unterschiedlichen Interessensschwerpunkt auf – in der vorliegenden Arbeit liegt der Fokus nicht auf dem Ubiquitous Computing. Die Ausarbeitung Kaminskis muss entsprechend nicht systematisch angepasst, sondern nur leicht kommentiert eingesetzt werden, um einen Mehrwert bieten zu können.

Ein Hauptteil dieses Mehrwerts besteht darin, mit Hilfe der Perspektive Foersters und den Weiterentwicklungen Kaminskis genauer zu klären, ob und wie der „Teil“ eines MLA, der die systematische Erstellung der Strukturvorschläge möglich macht, von eben diesen Strukturvorschlägen getrennt gedacht werden kann. Im Rahmen der Suche nach Problembegriffen in Abschnitt 3.4 wurde dieser Teil der Artefakte bereits mit dem Begriff der Vor-Struktur angenähert. Unter anderem in der Diskussion der transklassischen Technik wurde anschließend die Idee einer Vorstruktur, auf der die Erstellung des Strukturvorschlags basiert, wieder aufgegriffen. Die nachfolgende Darstellung der Perspektive Foersters beziehungsweise Kaminskis wird darauf aufbauend den Begriff der Vorstruktur konkreter an den Kontext des maschinellen Lernens anpassen und soll zeigen, dass die Diskussion verschiedener Vorstrukturen mitunter ebenso produktiv ist, wie die Unterscheidung zwischen zielorientierten und neugierigen maschinell lernenden Artefakten.

Foerster bestimmt **triviale Maschinen** als solche Maschinen, die auf die gleiche Eingabe immer auf die gleichen Art und Weise reagieren. Die darauf aufbauende Bestimmung einer **nichttrivialen Maschine** besteht darin, dass Technik betrachtet wird, die sich aus zwei **Maschinenkomponenten** zusammensetzt. Die eine Komponente stellt eine triviale Maschine der genannten Art dar, während die andere Maschinenkomponente die Transformationsprozesse – die Eingabe-Ausgabe-Zusammenhänge – adaptiert. Dieser zweite Vorgang wird als Manipulation der **Transformationsfunktion** des Artefaktes bezeichnet. Vor einer möglichen Übertragung von Foersters Unterscheidung zwischen trivialen und nichttrivialen Artefakten auf den Bereich des maschinellen Lernens stellt sich die Frage, inwiefern nichttriviale Artefakte als Technik betrachtet werden können, wenn sie auf dieselbe Eingabe unterschiedlich reagieren. Foerster selbst würde sagen, dass nichttriviale Maschinen im

¹³³ Als „smart“ werden, wie bereits beschrieben, Artefakte bezeichnet, die eine Nutzerabsicht antizipieren, bevor der Nutzer diese geäußert hat.

Grunde keine Technik sind, da für ihn alle Technik auf Trivialisierung abzielt. Diese These soll hier nicht diskutiert werden, es soll jedoch festgehalten sein, dass eine Betrachtung von Nichttrivialität per Konstruktion nicht ohne weitere Ausarbeitung geeignet ist, die im Vorigen gestellte Frage, ob MLA Technik sind, zu beantworten. Darüber hinaus besteht für den Einsatz der Perspektive im Kontext des maschinellen Lernens das Problem, dass für Foerster Informationen zu den allgemeinen Strukturen jedes Gegenstandsbereiches gehören (Kaminski 2012, S. 6). Hieraus ergeben sich gewisse begriffliche Unschärfen, wenn im Kontext von maschinellern Lernen von Informationen die Rede ist. Die Verwendung dieses Begriffes wird jedoch, wie bereits andiskutiert, sowieso vermieden, daher sei hier nur auf die Problematik hingewiesen. Kaminski diskutiert noch weitere Probleme, die bei der Nutzung des Konzeptes der Nichttrivialität auftreten, aber es genügt an dieser Stelle festzuhalten, dass der Begriff in seiner ursprünglichen Form im Kontext der Diskussion maschinellen Lernens nicht konstruktiv nutzbar ist. Beispielsweise ist in Foersters Begriffsbildung sehr vieles eine Maschine, je nach Auslegung sogar der Mensch, was viele in dieser Diskussion getroffene Aussagen nutzlos beziehungsweise redundant werden lässt (Kaminski 2012, S. 7). Weiter wird die Unterscheidung zwischen trivial und nichttrivial zu einfach angelegt, so dass einerseits alle Technik trivial ist und andererseits alle Technik nichttrivial ist, sofern sie stofflich ist. Der Grund hierfür ist, dass stoffliche Technik alterungsbedingt funktionsuntüchtig wird oder mangels ausreichendem Treibstoff ungewollt reagiert und so fort. Foersters Veranschaulichungen nichttrivialer Maschinen schließlich werden lediglich aufgrund ihrer großen Komplexität als nichttrivial bezeichnet – ein Beispiel hierfür ist der Chiffrierautomat.

Kaminski erarbeitet eine Weiterentwicklung des Konzeptes der Nichttrivialität, die diese Schwierigkeiten beseitigen soll und bereitet diese Weiterentwicklung sogar explizit auf einen Einsatz im Gebiet des maschinellen Lernens vor. Die Grundbeobachtung dabei ist, dass alle MLA nichttrivial im Sinne Foersters sind, beispielsweise wird ein KNN, das Backgammon spielt, je nach Fortschritt des Autoadaptionsprozesses unter identischen Rahmenbedingungen den Zug, den es zu einem früheren Zeitpunkt durchgeführt hat, nicht wiederholen. Die resultierende These Kaminskis ist, dass die bloße Rede von Nichttrivialität im Sinne Foersters ein zu simpel konzipierter Begriff ist, um lernende Algorithmen konstruktiv erfassen zu können. Kaminskis Bestreben ist, über den **Erwartungsbegriff** (Kaminski 2010) Ordnung und eine größere Präzision in den Begriff der Nichttrivialität zu bringen.

Je nach dem, in welchem Maße die Regeln einer Maschine erkennbar sind, lässt sie sich einem Niveau an Trivialität oder Nichttrivialität zuordnen. Die Differenzierung erfolgt dabei über den Modus der Transformationsfunktion.

(Kaminski 2012, S. 9)

Die Maschinenkomponente, die die Transformationsfunktion anpasst, entspricht im Kontext des maschinellen Lernens der Lernstrategie und ihrer Realisierung als eine **Vorstruktur**. Die Transformationsfunktion entspricht dem Strukturvorschlag als Resultat des Autoadaptionsprozesses. Der Strukturvorschlag ist von der zugrunde liegenden Lernstrategie in gewisser Hinsicht unabhängig und kann ohne Weiteres isoliert von ihr betrachtet werden. Die Idee Foersters zwei Maschinenkomponenten zu betrachten, ist daher sehr gut geeignet, das Verständnis maschinellen Lernens zu verbessern. Im Weiteren wird eine Modellierung beschrieben, die dieses Zusammenwirken zweier Maschinenkomponenten in den Fokus

nimmt¹³⁴. Der Autoadaptionsprozess umfasst in dieser Modellierung zwar sowohl die Vorstruktur als auch den Strukturvorschlag, allerdings legt die Vorstruktur fest, nach welchen Prinzipien die Adaptionen durchgeführt werden. Die Vorstruktur ist dabei aufgrund der Intransparenz oder Komplexität des Autoadaptionsprozesses durch den Nutzer gegebenenfalls nicht erkennbar, liegt jedoch immer vor. Weiter kann die Vorstruktur vorschreiben, dass sie selbst im Rahmen der Autoadaption fix bleibt oder auf welche Weise der Strukturvorschlag angepasst wird. Insbesondere kann die Vorstruktur sogar die Regeln adaptieren, nach denen sie selbst adaptiert wird. Im Rahmen mancher Autoadaptionsprozesse wird die Vorstruktur sogar insofern adaptiert, dass zukünftig andere Adaptionen der Vorstruktur möglich sind – die Regeln nach der sie selbst adaptiert wird, werden ebenfalls adaptiert. Anders formuliert, werden neben dem Vorwissen, auf dessen Basis der Strukturvorschlag entwickelt wird, auch die Regeln adaptiert, nach denen das Vorwissen und die Regeln geändert werden dürfen. Ein Beispiel hierfür ist die Adaption **evolutionärer Operatoren**.

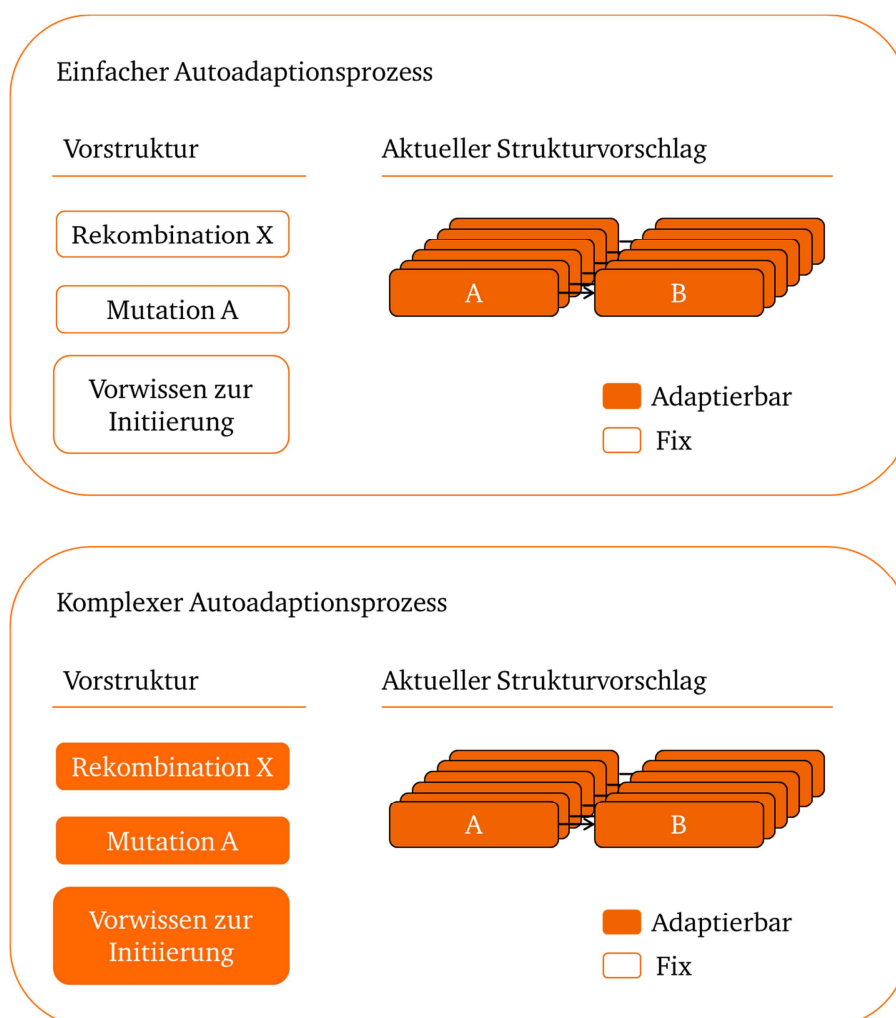


Abbildung 47: Einfache und komplexe Autoadaption im evolutionären Lernen

Diese Beispiele zeigen, dass die Aufteilung in zwei Maschinenkomponenten noch nicht ausreicht. Eine Modellierung von MLA auf Basis des Zusammenwirkens zweier Maschinenkomponenten steht noch immer vor der Herausforderung, verschiedene Typen von

¹³⁴ Kaminski verwendet die Begriffe etwas anders, worauf später noch eingegangen wird.

Nichttrivialität zu unterscheiden¹³⁵. Zur Erarbeitung solcher Typen von Nichttrivialität soll analog zu den Betrachtungen bei Kaminski angenommen werden, dass der Strukturvorschlag als triviale Artefaktkomponente Eingaben in Ausgaben umwandelt und dass die Transformationsfunktion gemäß der Vorstruktur als zweite Artefaktkomponente das MLA abhängig von den bisherigen Eingaben adaptiert. Auch im sehr häufig auftretenden Fall, dass die Vorstruktur selbst fix ist, kann maschinelles Lernen vielfältiger Art möglich sein. Der Autoadaptionsprozess kann etwa darin bestehen, bei jeder Übergabe eines speziellen Eingabedatums die Ausgabe pauschal um einen gewissen Wert zu erhöhen und bei jeder Übergabe eines zweiten festgelegten Eingabedatums die Ausgabe pauschal um denselben Wert abzusenken. Praktische Beispiele dieses sehr einfachen Verhaltens können häufig bei smarten Konsumgütern angetroffen werden. Eine konkrete Anwendung wäre die Anpassung der Anschlaggeschwindigkeit einer Tastatur in Abhängigkeit der vom Nutzer begangenen Fehler durch doppelte oder fehlende Zeichen. MLA dieser Art reagieren auf Eingaben, indem sie eine im Vorhinein festgelegte Adaption der Transformationsfunktion vornehmen. Dies wird bei Kaminski als **nichttriviale Nichttrivialität erster Ordnung** bezeichnet – allerdings wird bei Kaminski angenommen, dass alle maschinell lernenden Artefakte in die nachfolgende beschriebene Gruppe der nichttrivialen Nichttrivialität zweiter Ordnung fallen. Nichttriviale Nichttrivialität erster Ordnung beschreibt die Mehrzahl der in der Praxis eingesetzten MLA, wobei die Autoadaptionsprozesse jedoch meist so komplex sind, dass sie nicht ohne Weiteres analysiert oder verstanden werden können. **Nichttriviale Nichttrivialität zweiter Ordnung** betrifft Lernstrategien wie evolutionäres Lernen, bei denen die Adaptionsregeln der Vorstruktur wiederum Adaptionen unterworfen sind. Ein anderes Beispiel nichttrivialer Nichttrivialität zweiter Ordnung ist hypothesenbasiertes Lernen, wie es auch von Kaminski selbst angesprochen wird.

Dazu entwirft [die Lernstrategie] eine Hypothese für gegebene und validierte In- und Outputdaten, welche deren Ordnung modelliert. Werden nun weitere Daten gegeben, so wird die hypothetische Ordnung auf ihre Stimmigkeit geprüft und, falls Abweichungen auftreten, verändert, indem eine verfeinerte Ordnungshypothese gebildet wird, was den Lernprozess darstellt.

(Kaminski 2012, S. 12)

Nichttrivialität in Bezug auf Neugier und stabile Präsentationen

Nachdem der Bereich des maschinellen Lernens mit Hilfe der Nichttrivialität in die Teilbereiche der nichttrivialen Nichttrivialität erster und zweiter Ordnung unterteilt wurde, liegt die Frage nahe, ob und wie diese Unterteilung in Hinblick auf das Kontinuum zwischen zielorientierten und neugierigen MLA einzuordnen ist. Wenn nichttriviale Nichttrivialität als Maß der Veränderlichkeit der Komponenten des Autoadaptionsprozesses betrachtet wird, liegen mit den beiden Unterscheidungen vergleichbare Unterscheidungskriterien vor. Die Darstellung einer Zuordnung der in der Praxis eingesetzten MLA bezüglich dieser beiden Unterscheidungskriterien würde aller Voraussicht nach das folgende Bild entstehen lassen.

¹³⁵ Wobei auch nach Klärung dieses Punktes keine vollständige Auflösung der unterschiedlichen Lernstrategien möglich wäre. So wäre weiterhin offen, was etwa die Spezifik der KNN gegenüber dem evolutionären Lernen wäre. Beide Lernstrategien sind prinzipiell in der Lage, ihre Vorstruktur in nahezu beliebiger Weise zu adaptieren.

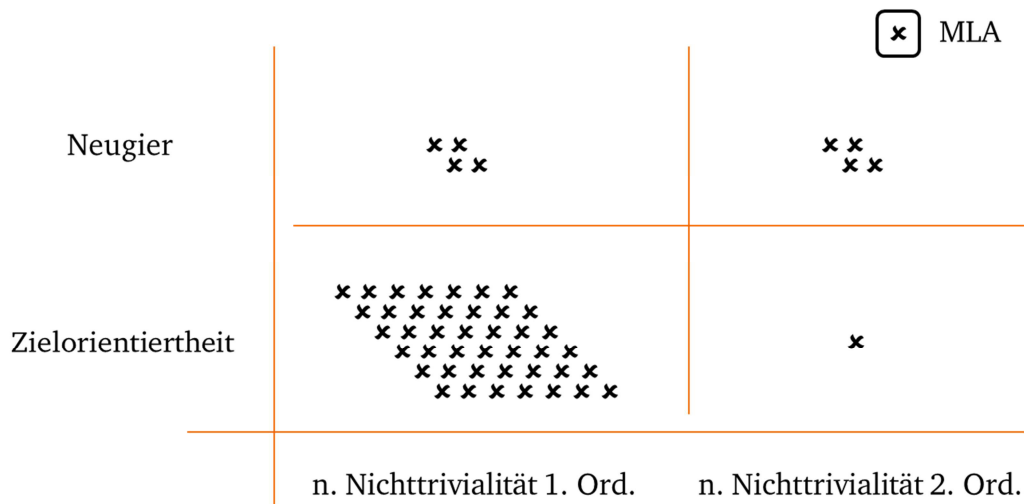


Abbildung 48: Geschätzte Häufigkeiten des Auftretens von MLA

Mit großer Wahrscheinlichkeit würde eine solche Auswertung ergeben, dass das Auftreten der beiden Kriterien für zielorientierte Artefakte korreliert, allerdings ist der Grund hierfür keine theoretische Notwendigkeit, sondern die bessere Optimierbarkeit von Systemen geringerer Komplexität. In Hinsicht auf neugierige MLA wäre die Gegenüberstellung der beiden Kriterien voraussichtlich nicht sehr aussagekräftig, weil die Erstellung von neugierigen Artefakten gerade nicht auf ein konkretes Ziel gerichtet ist. Aufgrund dessen kann spekuliert werden, dass die Wahl der Vorstruktur von den Rahmenbedingungen der jeweiligen Implementierung abhängig ist und eher selten gezielt getroffen wird. Die wesentliche Aussage dieser Spekulationen liegt in der These, dass es zunächst kaum Veranlassung gibt, die Konzepte der Neugier und der Nichttrivialität für deckungsgleich oder zu stark verwandt zu halten. Nichttrivialität bietet eine nützliche Perspektive, maschinelles Lernen generell zu betrachten und könnte das Potenzial haben, die Rede von der Selbstorganisation zu präzisieren. Neugier hingegen beschreibt einen Teilbereich des maschinellen Lernens, der technikphilosophisch von besonderem Interesse ist, da er eine systematisch neue Form von Technik zu sein scheint.

In der vorausgegangenen Abgrenzung des maschinellen Lernens vom Begriff der Repräsentation wurden Strukturvorschläge als stabile Visualisierungen eines sonst flüchtigen Blickes auf chaotische Rohdaten beschrieben. Höhlenmalereien wurden diesbezüglich als ein Beispiel für eine stabile Präsentation beziehungsweise Visualisierung außerhalb des maschinellen Lernens aufgeführt. Eine These lautete, dass das Konzept der Erwartung als eine Charakterisierung für den Umgang mit Technik höherstufig verwendet auch andersartige Technik noch zu modellieren erlaubt und dass MLA in gewisser Hinsicht die Funktion der Höhlenmalerei übernehmen¹³⁶. Das Besondere an maschinellem Lernen wäre in dieser Hinsicht, dass es diejenige Form von Höhlenmalerei darstellt, die den Nutzer in der Präsentationsfunktion vor unvorhergesehene Strukturen stellen kann. Genau hier könnte ein Anknüpfungspunkt zu Foerster bestehen, wenn die beiden Thesen vertreten würden, dass erstens Nichttrivialität eine Eigenschaft von Systemen ist und dass zweitens Trivialität und Nicht-Trivialität nicht eine Eigenschaft der Maschine, sondern das Verhältnis des Menschen zur Maschine kennzeichnen. In diesem Fall wäre die Höherstufigkeit mitgedacht, es ginge dann nicht mehr darum eine triviale Maschine herzustellen, sondern darum das Verhältnis zu

¹³⁶ Wobei festzuhalten ist, dass auch unter dieser These die objektstufige Verwendung von Überraschung, Erwartbarkeit und ähnlichen Konzepten, wie sie die klassische Steuer- und Regelungstechnik betreffen, zurückzuweisen ist.

trivialisieren. Ein Beispiel wären die sprachenlernenden MLA, bei denen ein derartiger Effekt auftritt. Das MLA ist in diesem Fall auf eine bestimmte Person oder eine spezielle Situation des Sprechens trainiert und unter anderen Rahmenbedingungen reagiert das Artefakt anders. Der genannte Effekt besteht nun darin, dass die Sprecher beginnen, sich an dieses Verhalten des Artefaktes zu adaptieren, sie verhalten sich nichttrivial zu dieser nichttrivialen Maschine und das Ziel besteht darin, wieder Trivialität zu erzeugen. Eine Möglichkeit, das Verhältnis zu einem MLA auf theoretischer Seite zu trivialisieren, besteht darin, ein vollständiges theoretisches Wissen zu erwerben, um alle ablaufenden Schritte des Autoadaptionsprozesses nachvollziehen und je nach Bedarf prognostizieren zu können. Die beiden genannten Thesen würden sich auch im Rahmen von Foersters Begriffsdifferenz zwischen Objekteigenschaften und Interaktionseigenschaften vertreten lassen. Die Begriffe triviale und nichttriviale Maschine wären in diesem Fall nicht an einem Objekt oder an Objekteigenschaften festgemachte Begriffe, sondern sie würden ein Verhältnis zu oder eine Interaktion mit einer Maschine kennzeichnen. Wenn die beiden Thesen zur Nichttrivialität des Verhältnisses nicht geteilt werden, wird es komplexer, von einer Höherstufigkeit zu sprechen. In diesem Fall ermöglicht der Begriff der Nichttrivialität noch immer die dargestellte Auflösung des maschinellen Lernens und ermöglicht so erst eine präzise Abgrenzung von etablierten Begriffsbildungen, die nicht pauschal das MLA oder den Autoadaptionsprozess betrachten. Das Konzept der Nichttrivialität kann entsprechend unabhängig vom Konzept der Höherstufigkeit und der Fassung als Eigenschaft von oder als Verhältnis zu Artefakten dabei unterstützen, diejenigen Algorithmen und Autoadaptionsprozesse zu identifizieren, die für die Technikphilosophie und insbesondere für die spätere Beschreibung einer Welttechnik relevant sein könnten. Dennoch ist besonders die zweite These durchaus interessant, da mit ihrer Hilfe ein besonderes Phänomen in den Blick rückt. Wenn etwa ein MLA Strategien des Schachspiels präsentiert und diese vom Nutzer repräsentiert werden, wird dieses MLA gerade dann besonders interessant, wenn es nicht nur gut funktioniert, wenn zwei Schachprogramme gegeneinander spielen, sondern wenn die Nutzer des MLA plötzlich gegen Bobby Fischer erfolgreich sind, aber gegen Garry Kasparov *nicht*. Die Eigenschaft, die in diesem Fall erlernt wurde, ist heuristisch gesehen das Schachspielen, aber damit erschöpft sich die Situation von der menschlichen Seite noch nicht, auch wenn das von der technischen Seite der Fall ist. In dieser Situation wäre die Nichttrivialität gewissermaßen immer schon höherstufig, nämlich reflexiv zu dem entsprechenden MLA als Mittel. Das wiederum würde die These ermöglichen und motivieren, dass diese Dimension nicht vernachlässigt werden darf. Das heißt, dass Teile des maschinellen Lernens als genau diejenige Technik identifiziert werden, bei der man von dieser Bezogenheit nicht abstrahieren kann. Dieser Gedanke soll hier nicht weiterverfolgt werden, stellt jedoch einen interessanten Anknüpfungspunkt an diese Analyse dar.

3.5.4. Fazit der drei Abgrenzungen

Sowohl die Unterscheidung zwischen klassischer und transklassischer Technik als auch diejenige zwischen naturalisierter und nicht-naturalisierter Technik stellen keine scharfen Disjunktionen dar. Zwischen den jeweiligen Begriffspaaren liegt noch sehr viel Niemandsland. Die im Vorigen beschriebene Angleichung beider Perspektiven konnte und sollte dieses Niemandsland lediglich etwas besser greifbar machen und somit den Suchraum für eine Verortung des maschinellen Lernens besser beschreibbar werden lassen. Gleichzeitig ist die Verortung neugierigen Lernens in diesem zweifach aufgespannten Niemandsland keineswegs endgültig oder zwingend, sondern die Argumentation könnte stattdessen auch weitergehend

betrachten, ob neugieriges maschinelles Lernen weder klassischen noch transklassischen, sondern zum Beispiel künstlerischen Charakter hat. Die Argumentation in diesem Zusammenhang könnte lauten, dass die Nutzer ein Ding vorfinden, das irgendeinen opaken Prozess durchläuft, den wir formal Autoadaptionsprozess nennen, und dessen Ergebnis vom Nutzer als ästhetisch ansprechend wahrgenommen wird oder nicht, wie es bei Brown (Brown et al. 2007) umgesetzt ist¹³⁷. In der Betrachtung der Erzeugung von Welt in Abschnitt 3.6 wird dieser Gedanke noch einmal aufgegriffen.

Aus der Betrachtung der transklassischen Technik wurde die Idee gewonnen, dass mit neugierigen Artefakten höherstufig wieder eine klassische Form von Technik vorliegt, da der Nutzer gezwungen ist, sich mit den Strukturvorschlägen auseinander zu setzen. Die Betrachtung der Nichttrivialität maschinell lernender Artefakte zeigte weiterhin, dass in Bezug auf den Autoadaptionsprozess mitunter keine „stabile“ Technik vorliegt, da die Transformationsfunktion nicht fix sein muss, dass aber die Bedingungen, die den Grad der Instabilität der Transformationsfunktion festlegen, durchaus kontrollierbar oder steuerbar sind beziehungsweise sein können. In einem solchen Fall läge die Stabilität neugieriger und zielorientierter Artefakte in der Eigenschaft, Unvorhergesehenes in Strukturvorschlägen betrachtbar zu machen, während für zielorientierte Artefakte zusätzlich eine Stabilität des Weges zu diesen Strukturvorschlägen einforderbar wäre. Die Sicherstellung dieser Stabilität des Weges zum Strukturvorschlag ließe sich gegebenenfalls als eine höherstufige Art der Steuerung beziehungsweise Regelung interpretieren, wobei die genauen Ausformulierungen dieser Perspektive einen weiteren Anknüpfungspunkt zu der vorliegenden Arbeit darstellen. An dieser Stelle wird auf die Ausarbeitung zielorientierter MLA zu Gunsten einer Konzentration auf neugieriges maschinelles Lernen verzichtet. Festgehalten werden soll jedoch, dass sowohl die Perspektive der transklassischen Technik als auch diejenige der nichttrivialen Technik die Intuition stärken, zielorientiertes maschinelles Lernen höherstufig als klassische Technik zu betrachten. Gleichwohl ist die Unterscheidung von nichttrivialer Nichttrivialität erster und zweiter Ordnung nicht mit der Unterscheidung zwischen neugieriger und zielorientierter Technik gleichzusetzen. Die zweite Perspektive beschreibt eine Eigenschaft des MLA selbst, während die erste immer auch ein Verhältnis des Nutzers zur Technik im Blick hat. Bayes'sches Lernen etwa kann neugierig sein, die Adaptionsregeln des Autoadaptionsprozesses sind jedoch meist fix. Eine heuristisch zu prüfende Annahme wäre, dass dieses Zusammentreffen nicht typisch ist, da die Adaptivität der Vorstruktur die Komplexität des MLA erhöht und die Vorhersehbarkeit der Strukturvorschläge senkt. Die zu prüfende These wäre, dass wachsendes Hintergrundwissen bezüglich der Arbeitsweise des Autoadaptionsprozesses sowie der Herkunft und den Rahmenbedingungen der Rohdaten den Nutzer dazu tendieren lassen, ein neugieriges MLA auf autoadaptiven Vorstrukturen basierend zu entwerfen.

Insgesamt wurde in dieser Abgrenzung von aktuellen technikphilosophischen Perspektiven versucht zu identifizieren, in welchen Teilbereichen von Technik maschinelles Lernen verortet werden kann. Ein Ziel dabei war zu zeigen, mit welchen Arten von Technik sich maschinelles Lernen besonders gut vergleichen lässt. Die Diskussion des Konzeptes von Selbstorganisation betrifft nicht nur das maschinelle Lernen. Ein Ziel dieser Arbeit ist es, die Diskussion maschinellen Lernens soweit zu präzisieren, dass sie in Kontexten, die nicht nur maschinelles Lernen betreffen, als Quelle hilfreicher und sachlich korrekter Intuitionen und Beispiele

¹³⁷ In diesem Fall könnte sogar auf die nutzerseitige Interpretation verzichtet werden.

fungieren kann. Eine später noch weiter ausgeführte Vermutung ist in diesem Zusammenhang, dass die durch maschinelles Lernen exemplifizierte Form von Technik nicht nur in Teilbereichen der Informatik, sondern auch in Technikbereichen, die aus den Nano-, Bio- und Neurowissenschaften resultieren, auffindbar ist. Die Vermutung, dass in anderen Technikbereichen Selbstorganisationsprozesse im Sinne der hier beschriebenen Ansätze der neugierigen Technik zum Einsatz kommen, erscheint gerade für die Biotechnik besonders plausibel und eine daran anknüpfende Forschung wirkt besonders aussichtsreich. In der Nanotechnik hingegen werden vorwiegend Umgebungen stimuliert und zu manipuliert, um eine Form von medialer Steuerung zu etablieren (Wiegerling 2012) – aber auch hier ist die Diskussion neugieriger Technik vielversprechend. Zusammengefasst besteht die These darin, dass nicht nur der Begriff der zielorientierten Technik, sondern auch derjenige der neugierigen Technik sich auf Kontexte jenseits des maschinellen Lernens ausweiten lässt.

3.6. Erzeugung von Welt und maschinelles Lernen

Die vorausgegangene Betrachtung technikphilosophischer Entwürfe diente der Klärung, wie eine Beschreibung von maschinellern Lernen gelingen kann und welche Aspekte und Strukturmomente von MLA einer besonderen Betrachtung bedürfen. Anknüpfungspunkte für eine Diskussion zielorientierter MLA wurden benannt und es wurde aufgezeigt, dass bei der präzisen Aufarbeitung neugieriger MLA aus Sicht der Technikphilosophie mehr Neuland betreten wird als bei der Beschäftigung mit den in der Praxis oft stark mathematisch geprägten Optimierungsalgorithmen des zielorientierten Lernens. Im Weiteren wird der Fokus dementsprechend speziell auf neugierigem maschinellern Lernen und den resultierenden unvorhergesehenen Strukturvorschlägen liegen. Die erste Nutzung der Grundintuition, dass die Rede von Welt im Kontext des maschinellen Lernens einen Mehrwert bietet, fand in Abschnitt 3.4 statt und diente dazu, den Blick zu verbreitern um zusätzliche Fragestellungen zu identifizieren. Die folgende Betrachtung eines zweiten Weltbegriffes hingegen soll die Analyse wieder fokussieren¹³⁸.

Die Absicht hinter der folgenden Betrachtung der Modellierungen in Goodmans *Ways of Worldmaking* liegt insbesondere darin, die Rede von Welttechnik vorzubereiten. Vor diesem Hintergrund ist von besonderem Interesse, dass Goodman mit den *Weltversionen* ein Konzept vorstellt, das unabhängig von maschinellern Lernen entwickelt und diskutiert wird, jedoch große Parallelen dazu aufweist.

[Modern philosophy] exchanged the structure of the world for the structure of the mind, [...] exchanged the structure of the mind for the structure of the concepts, and [...] proceeds to exchange the structure of concepts for the structure of [...] perception. [...] The movement is from unique truth and a world fixed and found to a diversity of right and even conflicting versions or worlds in the making.

(Goodman 1978, Vorwort S. "x")

Attention usually focuses on versions that are literal, denotational, and verbal. While that covers some [...] scientific [...] worldmaking, it leaves out perceptual and pictorial versions and all figurative and exemplificational means and all nonverbal media.

[...]

Such worldmaking and such versions are my primary concern here;

(Goodman 1978, S. 102)

Ein Strukturvorschlag stellt genau dann eine solche Weltversion dar, wenn er an einen Nutzer übergeben wird – wenn er an ein technisches System übergeben wird, bleibt er ein Strukturvorschlag. Der Begriff der Weltversion bezeichnet die verschiedenen Weisen Strukturen zu präsentieren, und Welt entspricht der Wahrnehmung der Strukturen, die durch die Weltversionen erzeugt werden. Diese Trennung ist eigentlich nicht strikt, da ein Künstler

¹³⁸ Die Verbreiterung des Blickes diente nicht nur der Gewinnung von Fragestellungen und Anknüpfungspunkten, sondern auch der Nutzbarmachung technikphilosophischer Entwürfe, die zum Teil nicht explizit in Hinblick auf maschinelles Lernen formuliert wurden.

die entsprechende Welt gegebenenfalls bereits mitgeschaffen haben kann. Allerdings ist dieser Fall für die Betrachtung von MLA nicht von Bedeutung, wie etwa in der Diskussion des gestifteten Worumwillen dargestellt wurde. Insgesamt wird in der nachfolgenden Betrachtung diskutiert, ob und wie neugierige Artefakte Menschen bei der Erstellung von Weltbezügen, wie Goodman sie in *Ways of Worldmaking* entwirft, unterstützen können.

Die Analyse der Möglichkeiten und Wege, wie MLA den Menschen unterstützen können, basiert auf derselben Frage, die auch Goodmans Ausarbeitung zugrunde liegt.

We face the question how worlds are made, tested, and known.

(Goodman 1978, S. 7)

My approach is [...] an analytic study of types and functions of symbols and symbol systems.

(Goodman 1978, S. 5)

Die Erstellung von Strukturvorschlägen, die **symbolische Systeme** darstellen, ist entsprechend Teil von Goodmans Analyse und auch die Betrachtung von subsymbolischen Strukturvorschlägen oder einer subsymbolischen Vorstruktur stellt kein prinzipielles Problem dar. Generell kann jeder Autoadaptionsprozess, das heißt jede Form der Raffinierung von Rohdaten zur Erstellung von Strukturvorschlägen, ohne Schwierigkeiten im Rahmen von Goodmans Überlegungen mitbetrachtet und analysiert werden. Die Frage, ob und in welcher Hinsicht die an den Menschen übermittelten Strukturvorschläge diesen bei der Erzeugung von Welten unterstützen können, kann entsprechend aus dieser Perspektive sinnvoll gestellt werden. Insbesondere die von Goodman beschriebene Rolle des Wissenschaftlers bei der Beobachtung von Messwerten hat eine interessante Schnittmenge mit der Funktionalität eines Autoadaptionsprozesses.

[The scientist] looks to the multifaceted and irregular results of observations for little more than suggestions of overall structures and significant generalizations. He seeks system, simplicity, scope; and when satisfied on these scores he tailors truth to fit [...]

[...]

World are made [...] also by what is exemplified and expressed – by what is shown as well as by what is said.

(Goodman 1978, S. 18)

Der Einsatz von maschinellem Lernen lässt sich als eine Auslagerung von Teilen der Beobachtung modellieren, bei der anstelle eines menschlichen Beobachters ein maschinell lernendes Artefakt aktiv ist und die Strukturvorschläge anschließend auf die von Goodman genannte Weise von den Nutzern aufgenommen werden.

Genau wie bei Goodman wird auch in der Diskussion des maschinellen Lernens keine vollständige Antwort auf die Frage nach der Wahrheit oder überhaupt der Qualität der

entstandenen Welt benötigt. Das Ziel der Aufarbeitung von Goodmans Perspektive liegt darin, die denkbaren Ansatzpunkte einer technischen Unterstützung bei der Welterzeugung zu benennen und in Bezug auf MLA zu diskutieren. Entsprechend wird Goodmans Vorhaben einer Kritik der Welterzeugung (Goodman 1978, S. 94) hier nur in Ansätzen relevant. Der Fokus wird vielmehr darauf liegen, überzeugend darzulegen, dass maschinell lernende Artefakte technische Unterstützung bei der Welterzeugung anbieten – darüber hinaus gehende Fragen bieten sich als Anknüpfungspunkte für weitere Projekte an.

3.6.1. Parallelen des maschinellen Lernens zur Kunst

Die Möglichkeit eines Vergleichs zwischen Strukturvorschlägen und Kunstwerken wurde in der Abgrenzung des maschinellen Lernens von der Informationstechnik und vom Begriff der Repräsentation, sowie bei der ersten Betrachtung des Weltbegriffes bereits angedeutet¹³⁹. Diese Betrachtung wird nachfolgend noch etwas ausgebaut, da Kunst auch in der Analyse Goodmans thematisiert wird und dort ein wesentliches Hilfsmittel für den Erkenntnisgewinn darstellt.

[...] a major thesis of this book is that the arts must be taken no less serious than the sciences as modes of advancement of the understanding [...].

(Goodman 1978, S. 102)

Goodman begründet zu Beginn seiner Analyse, dass Kunst der aufnehmenden Person eine Möglichkeit bietet, die eigene Mustererkennung von außen inspirieren zu lassen.

The notational system [of two musical performances of the same work] distinguishes constitutive from contingent features, thus picking out the performance-kinds that count as works [...] “now I can go”, in Wittgenstein's sense, when I have found a familiar pattern, or a tolerable variation of one, that fits and goes beyond the cases given.

(Goodman 1978, S. 10)

Diese Eigenschaft beschreibt genau dasjenige, was maschinelles Lernen für den Nutzer leisten soll. Goodmans Darstellung lässt sich direkt auf einen Autoadaptionsprozess übertragen, der ausgehend von Trainingsdaten Strukturvorschläge erarbeitet. Neugieriges maschinelles Lernen kann Muster in einer Aufführung eines musikalischen Werkes auch dann noch identifizieren, wenn die zugrunde liegenden musikalischen Werke im Vorhinein nicht bekannt sind. In der bisherigen Analyse wurde ausgearbeitet, dass neugierig lernende Artefakte Strukturvorschläge genannte Systeme von Beschreibungen entwerfen, die vom Nutzer interpretiert werden, um den Rohdaten einen Referenzrahmen geben zu können. Jeder Strukturvorschlag stellt eine Systematik der Beschreibung der Rohdaten dar. Die Bedeutung dieser Art von Beschreibungssystematiken betont Goodman explizit, wobei er sich auf die Werke von Künstlern konzentriert und diskutiert, inwiefern diese Werke Strukturvorschläge

¹³⁹ Weiter wurde mit (Brown et al. 2007) ein Beispiel genannt, in dem maschinelles Lernen unmittelbar als Kunst eingesetzt wurde.

darstellen, die in Kombination mit individuellen persönlichen Hintergründen eine große Vielfalt an Interpretationen ermöglichen.

Frames of reference, though, seem to belong less to what is described than to systems of description [...]

[...]

Much more striking is the vast variety of versions and visions in the several sciences, in the works of different painters and writers and in our perceptions as informed by these, by circumstances, and by our own insights, interests, and past experiences.

(Goodman 1978, S. 2f)

Die hier von Goodman betrachtete Wirkung von Kunst auf die Welterzeugung lässt sich gut auf den Einsatz von neugierigen MLA übertragen – auch diese lassen in der Wahrnehmung der Nutzer in Zusammenwirkung mit deren eigenen Einsichten, Interessen und Erfahrungen Neues entstehen. Die Parallele zwischen Strukturvorschlägen und Kunstwerken wird im Folgenden in der Diskussion des *Weighting* noch einmal aufgegriffen, jedoch nicht wesentlich weiterentwickelt und stellt damit einen möglichen Anknüpfungspunkt an diese Arbeit und eine offene Frage dar.

Unabhängig davon lässt sich jedoch festhalten, dass Goodmans Argumentation in Hinblick auf Werke der Kunst und deren Auswirkung auf den Menschen eine vergleichbare Nutzung seines Weltbegriffs in Hinblick auf das maschinelle Lernen und das Verhältnis zwischen Strukturvorschlägen und Menschen durchführbar und sinnvoll erscheinen lässt. Diese Übertragbarkeit der Rolle, die Goodman den künstlerischen Werken zuschreibt, auf die Strukturvorschläge maschinellen Lernens wird im Weiteren anhand der von ihm exemplarisch dargestellten Weisen der Welterzeugung weiter ausgearbeitet.

3.6.2. Weisen der Welterzeugung

Das Ziel dieser Aufarbeitung von Goodmans Perspektive liegt wie beschrieben darin, die denkbaren Ansatzpunkte einer technischen Unterstützung bei der Welterzeugung zu benennen und in Bezug auf MLA zu diskutieren. Einige der wichtigsten Weisen von Welterzeugung werden von Goodman konkret thematisiert und als Konstruktionsvorgänge beschrieben.

Whatever else may be said of these modes of organization, they are not “found in the world” but build into a world. Ordering, as well as composition and decomposition and weighting of wholes and kinds, participate in worldmaking.

(Goodman 1978, S. 14)

Diese aktive Konstruktion stellt einen Prozess dar, der von maschinellem Lernen technisch unterstützt werden kann. Die These, dass das Vorliegen eines solchen Konstruktionsprozesses nicht fraglich ist, erhöht dabei die Relevanz der Analyse der Möglichkeiten einer technischen Unterstützung eben jenes Konstruktionsprozesses. Die weitere Analyse dieser Möglichkeiten

wird demnach kompakt anhand der von Goodman genannten Beispiele für Weisen der Welterzeugung erfolgen.

(a) Composition and Decomposition

Das erste Beispiel Goodmans für eine Weise der Welterzeugung beschreibt die Erstellung und Auflösung von Klassen beziehungsweise Klassifizierungen.

(a) Composition and Decomposition

Much [...] worldmaking consists of [...] dividing into part [...], analyzing complexes into component features, drawing distinctions; on the other hand, of composing wholes and kinds out of parts [...], combining features into complexes, and making connections. Such composition or decomposition is normally effected or assisted or consolidated by the application of labels [...].

(Goodman 1978, S. 7)

Diese Vorgehensweise zeichnet sehr viele maschinelle Autoadaptionsprozesse aus und ist insbesondere zentraler Bestandteil aller Autoadaptionsprozesse, die einen Klassifikator erzeugen sollen. Goodman setzt hier nicht zwingend voraus, dass die Ausgangsbasis der Komposition und Dekomposition mit Hilfe von Messungen geschaffen werden muss, sondern auch der Einsatz von Vorwissen ist mitgedacht. Demgegenüber wird die Möglichkeit, dass die Ausgangsbasis komplett neu geschaffen wird, zumindest implizit verneint. *Worldmaking* wird als eine Art und Weise betrachtet, Weltversionen aus bereits bestehenden Weltversionen zu gewinnen¹⁴⁰. Das umschaffende *Worldmaking* als die Erstellung von Welten wird bei Goodman von einem erschaffenden *World-Building*, im Sinne einer Neuschöpfung von Welten, unterschieden.

Worldmaking as we know it always starts from worlds already on hand; the making is remaking.

(Goodman 1978, S. 6)

Goodman geht zwar davon aus, dass noch weitere, völlig andere Weisen der Welterzeugung möglich sind, die Frage nach der Möglichkeit einer Erschaffung einer prinzipiell neuen Welt würde er jedoch zumindest für Artefakte eher verneinen. Eine technische Unterstützung würde sich entsprechend auf die umschaffende Welterzeugung konzentrieren und könnte zwar auch ohne die Nutzung von MLA erfolgen, allerdings müsste jedes technisch unterstützende Artefakt, das auf Eingabedaten verzichtet, mit umfangreichem Vorwissen versorgt werden.

Die Rede von einer technischen Unterstützung des Welterzeugungsprozesses impliziert die Forderung, dass der Erzeugungsprozess sich aufteilen lässt, so dass ein MLA einen Teilschritt übernehmen kann. Dies korrespondiert mit der These, dass maschinell lernende Artefakte an sich noch keine Weltbezüge oder gar Welten generieren und dass erst der Mensch die dafür

¹⁴⁰ Goodman unterscheidet nicht strikt zwischen Welten und Weltversionen.

notwendige Interpretation vornimmt. Der Teilschritt des MLA besteht in der reinen Neuordnung oder Erstordnung von Rohdaten und reicht noch nicht aus um einen Weltbezug zu erstellen. Dieser Punkt wird von Goodman insofern explizit angesprochen, als er auf die Frage eingeht, ob die Interpretation von Strukturvorschlägen scheitern kann. Er betrachtet dabei die Frage, ob die reine Wahrnehmung eines Strukturvorschlages – beim Scheitern der Interpretation – auch *nicht* zu dem Entstehen eines Weltbezuges führen kann.

We do not make a new world every time we take things apart or put them together in another way; but worlds may differ in that not everything belonging to one belongs to the other.

(Goodman 1978, S. 8)

Sogar die Überanpassung an Rohdaten als ein Scheitern daran, eine hilfreiche Anzahl von Kategorien zu erstellen, wird von Goodman implizit erwähnt.

A world may be unmanageably heterogeneous or unbearably monotonous according to how events are sorted into kinds.

(Goodman 1978, S. 9)

Insgesamt werden Komposition und Dekomposition als Weisen der Welterzeugung im maschinellen Lernen umfassend abgebildet.

(b) Weighting

Das zweite Beispiel Goodmans für eine Weise der Welterzeugung beschreibt die Art wie Klassen in ihrer Bedeutung oder Wichtigkeit bewertet werden.

[(b) Weighting]

What counts as emphasis, of course, is departure from the relative prominence accorded to the several features in the current world of our everyday seeing.

[...]

These differences in emphasis, too, amount to a difference in relevant kinds recognized.

(Goodman 1978, S. 11)

Zwar erlaubt maschinelles Lernen unterschiedliche Wahrnehmungen von Rohdaten, Relevanzbestimmungen jedoch finden im Autoadaptionsprozess selbst nicht statt, sondern sind diesem vor- oder nachgelagert. Unterschiedliche Bestimmungen der Relevanz von Gruppen von Rohdaten können durch neugieriges Lernen ermöglicht, aber nicht erzeugt werden. Diese Bestimmungen basieren auf den ebenfalls von Goodman erwähnten Interessen der Nutzer, die wiederum Teil des Hintergrundwissens sind, das benötigt wird, um neugierig

entstandene Strukturvorschläge betrachten oder eine zielorientierte Lernstrategie entwickeln zu können.

Unabhängig von den Bestimmungen der Relevanz lässt sich an dieser Weise der Welterzeugung gut der Hauptunterschied zwischen Goodmans Rede von **Kunst** und der Diskussion des maschinellen Lernens erkennen.

The distinction between saying or representing on the one hand and showing or exemplifying on the other becomes even more evident in the case of abstract painting and music and dance [...].

(Goodman 1978, S. 12)

[...] mere possession of a property does not amount to exemplification [and] exemplification involves reference by what possesses to the property possessed [...].

(Goodman 1978, S. 32)

Die Strukturvorschläge neugieriger Algorithmen werden nicht von Künstlern erstellt, sondern zeigen oder präsentieren nur etwas. Künstler hingegen erstellen meist gerade keine großen Mengen anspruchloser Werke mit dem Wunsch, dass gegebenenfalls eines der Werke eine Verwend- oder Interpretierbarkeit aufweisen möge. Von einem Kunstwerk wird erwartet, dass es gewisse Eigenschaften exemplifiziert, dass es gezielt Eindrücke erweckt und meist auch, dass diese Eindrücke vom Künstler im Vorfeld prognostiziert und manipuliert werden. Der erwünschte Effekt beim Betrachtenden ist sowohl bei der Betrachtung eines Strukturvorschlages als auch der eines Kunstwerkes die Entdeckung einer neuen Perspektive, allerdings unterscheidet der Prozess sich dennoch deutlich. Dieser Bruch zwischen Kunst und maschinell Lernen nimmt jedoch auch aus Goodmans Perspektive der hier vorgenommenen Analyse maschinellen Lernens nicht ihre Relevanz, da auch nichtrepräsentative Werke denselben Zweck wie Exemplifikationen übernehmen können.

Nondescriptive, nonrepresentational works nevertheless function as symbols for features they possess either literally or metaphorically. Serving as samples of, and thereby focusing attention upon, certain – often upon unnoticed or neglected – shared or shareable forms, colors, feelings, such works induce reorganization of our accustomed world in accordance with these features, thus dividing and combining erstwhile relevant kinds, adding and subtracting, effecting new discriminations and integrations, reordering priorities.

(Goodman 1978, S. 105)

Dasselbe, was Werke für bekannte Zusammenhänge erreichen können, gelingt mitunter mit Hilfe von maschinell lernenden Artefakten für ungreifbare Rohdaten, die keine *Accustomed World* darstellen. Die Algorithmen **induzieren** entsprechend in der Wahrnehmung des Nutzers keine Re-Organisation, sondern eine initiale Organisation. Allenfalls stößt man auf das bereits diskutierte Problem, dass eine technisch erstellte Transformationsfunktion je nach Begriffsbildung erst dann zu einem *Sample* wird, wenn ein Nutzer sie als solche erkennt.

(c) Ordering

Das dritte Beispiel Goodmans für eine Weise der Welterzeugung beschreibt die Art wie die Zusammenhänge zwischen Klassen von Rohdaten beschrieben werden können.

[(c) Ordering]

Much as the nature of shapes changes under different geometries, so do perceived patterns change under different orderings;

[...]

All measurement [...] is based upon order. Indeed only through suitable arrangements and groupings can we handle vast quantities of material perceptually or cognitively.

(Goodman 1978, S. 13f)

Dieses Beispiel ist insbesondere für Data Mining von Interesse, bei dem die Rohdaten meist nicht klassifiziert werden sollen, sondern nach Abhängigkeiten gesucht wird. Während in den bisherigen beiden Beispielen für Weisen von Welterzeugung Klassifizierungen von Rohdaten im Fokus standen, sind hier alle Rohdaten von gleicher Bedeutung. Dies ist etwa dann der Fall, wenn Verkaufsdaten betrachtet werden und jeder Verkaufsvorgang von gleicher Bedeutung ist. Zwar sind die Fragen nach unterschiedlichen Klassen und die nach Ordnungen – und damit nach in einer Struktur verbundenen Daten – verwandt, aber die Fragestellungen haben eine ausreichend unterschiedliche Ausrichtung, um eine gesonderte Erwähnung zu rechtfertigen.

(d) Deletion and Supplementation

Diese Weise der Welterzeugung legt den Fokus darauf, dass Menschen bei der Erzeugung von Weltbezügen die tatsächlich zugänglichen Datengrundlagen verändern, indem sie vorliegende Rohdaten ignorieren oder unvollständige Daten nach eigenem Gutdünken ergänzen. Sowohl das Konzept der Auslassungen als auch das der Ergänzung lassen sich ohne Weiteres technisch realisieren. Ergänzung entspricht etwa – übertragen auf zielorientiertes maschinelles Lernen – Extrapolations- und Interpolationsprozessen, bei denen eine unzureichende Datenbasis mit Hilfe von vorgegebenen mathematischen Funktionen erweitert wird. Auslassungen treten sowohl im Kontext von neugierigem als auch von zielorientiertem Lernen auf und entsprechen beispielsweise der gezielten Vernachlässigung von Rohdaten oder von Vorwissen. Diese Vorgehensweise wird, wie im ersten Hauptteil beschrieben, vor allem bei der Bekämpfung von Überanpassung eingesetzt, etwa bei der Stützung von Entscheidungsbäumen. Derartige Maßnahmen gegen Überanpassung sind auch im Rahmen einer technischen Unterstützung bei der Erzeugung von Weltbezügen vergleichsweise unkritisch und sinnvoll, da Nutzer es aus vielen Kontexten gewohnt sind, mit Konzeptualisierungen und Modellierungen umzugehen, die nicht jedes Detail korrekt darstellen oder die auf mitunter sehr groben Annahmen beruhen.

(e) Deformation

Dieses letzte Beispiel für eine Weise der Welterzeugung beschreibt den Umstand, dass entstandene Weltbezüge nachträglich gezielt umgeformt werden. Diese Vorgehensweise wird

technisch im Rahmen des maschinellen Lernens ebenfalls realisiert, indem Strukturvorschläge entwickelt werden, die nicht den gegebenen Rohdaten entsprechen. Dies kann beispielsweise sinnvoll sein, wenn angenommen wird, dass ein Rauschen vorliegt und die Eingabedaten entsprechend unzuverlässig sind oder wenn ein systematischer Fehler des Autoadaptionsprozesses ausgeglichen werden soll. Umformungen von Weltbezügen können zielorientiert unterstützt werden, indem Rohdaten systematisch uminterpretiert werden, um etwa einen Bias in der Messtechnik oder der Vorstruktur auszugleichen. Alternativ kann die Umformung neugierig unterstützt werden, indem zur Reduzierung der Überanpassung des resultierenden Strukturvorschlages kleine, unsystematische Abwandlungen in die Daten eingebaut werden. In der Praxis sind beide Varianten der technischen Unterstützung üblich, um dem Nutzer die Entscheidung zu erleichtern, wie stark Rohdaten abgewandelt werden müssen, damit die aus den Strukturvorschlägen gezogenen Erkenntnisse eine über die konkret vorliegenden Rohdaten hinausgehende Aussagekraft haben. Insgesamt ist eine technische Unterstützung des Menschen durch MLA auch bei Umformungen als Weise der Welterzeugung ohne größere Schwierigkeiten möglich.

Die Bedeutung technischer Unterstützung der Welterzeugung

Im Vorherigen wurde dargelegt, dass maschinelles Lernen den Menschen bei der Welterzeugung technisch unterstützen kann, indem Strukturen präsentiert werden, die nach einer Interpretation eine Weltversion exemplifizieren. Strukturvorschläge dieser Art werden bei Goodman implizit mitgedacht und zugelassen.

We saw earlier that works or other symbols that do not declare or describe or represent anything, literally or metaphorically, or even purport to denote anything, may present worlds by exemplification.

(Goodman 1978, S. 133)

Diese Formulierung lässt zunächst die Möglichkeit offen, dass Strukturvorschläge Exemplifikationen darstellen müssen, um in Goodmans Perspektive zulässig zu sein. Dies kann jedoch verworfen werden, da die Notwendigkeit, einen Strukturvorschlag interpretieren zu müssen, von Goodman explizit mitgedacht und in seine Perspektive integriert ist.

Thus some samples and the nonverbal labels or features exemplified by or projectible from them may, unlike evidence statements and hypotheses, belong to symbol systems that are neither denotational nor articulate.

(Goodman 1978, S. 136)

Die nächste Frage ist, ob es sich bei der Unterstützung um einen relevanten Beitrag handelt. Goodman selbst äußert sich zum Gewinn von Wissen und Verständnis wie folgt.

An increase in acuity of insight or in range of comprehension, rather than a change in belief, occurs when we [...] study [...] until we see or hear or grasp features and structures we could not discern before. Such growth in knowledge is not by formation or fixation or belief but by the advancement of understanding

[...]

*Recognizing patterns is very much a matter of inventing and imposing them.
Comprehension and creation go on together.*

(Goodman 1978, S. 21f)

Die hier beschriebene Fähigkeit, unvorhergesehene Strukturen erkennen zu können, profitiert direkt von Strukturvorschlägen, wie sie von MLA erstellt werden können. Wie bereits ausführlich dargestellt, entwickeln neugierige MLA ihr Strukturvorschläge unabhängig von Nutzern und Entwicklern, während zielorientierte MLA zumindest in Teilen unabhängig vom Nutzer unvorhergesehene Strukturen entdecken oder entwickeln können. Das bedeutet, autoadaptive Artefakte können eine zentrale Aufgabe in dem von Goodman skizzierten Prozess der Gewinnung von Verständnis und Wissen übernehmen. Gerade diese Aufgabe stellt anderenfalls einen der wesentlichsten Engpässe in kreativen Prozessen dar.

The perception of any pattern not fitting the structure of the search often takes great trouble.

(Goodman 1978, S. 40)

Vor allem neugierige MLA sind der Lage, unvoreingenommen und autonom zu beobachten. Die Aufgabe, die Struktur der Suche von den Suchergebnissen zu trennen, kann an die Informatik überantwortet werden und ein dort entwickelter Lösungsansatz liegt im neugierigen maschinellen Lernen. So ist es möglich, die Suche nach Strukturen von der Interpretation der Strukturen zu trennen und beide Schritte getrennt zu analysieren und gegebenenfalls zu verbessern. Goodman selbst argumentiert ebenfalls dafür, dass zufällige und damit nicht vom Menschen verursachte Effekte dies leisten können.

Nothing here is incompatible with the familiar fact that interesting qualities are sometimes revealed through the juxtaposition of works in a mixed anthology, exhibition, or concert, or even a storeroom jumble.

(Goodman 1978, S. 39)

Das heißt, die Konstruktion eines technischen Hilfsmittels zur Unterstützung der neugierigen Suche nach unvorhergesehenen Strukturen ist aus Goodmans Perspektive durchaus denkbar. Weiter ähnelt seine Beschreibung menschlichen Sehens der Beschreibung eines Autoadaptionsprozesses.

The visual system drives toward uniformity and continuity, constrained by its anatomy and physiology, and influenced by what it has seen and done before, but improvising along the way.

(Goodman 1978, S. 79)

Diese Vorgehensweise findet sich auch bei den Ansätzen des maschinellen Lernens – wenngleich die Fähigkeit zur Improvisation je nach Denkweise nicht in Gänze durch technisch realisierte Autoadaptionsprozesse abgebildet werden kann. Das bedeutet, maschinelles Lernen ist als technische Unterstützung oder sogar Alternative zu einem menschlichen Sinn beschreibbar. Weiter wird diese Unterstützung gerade nicht von zielorientierten, an den Kontext angepassten MLA geleistet, die – etwa im Rahmen einer Verkehrszeichenerkennung – bereits interpretierte Konzepte übermitteln, sondern von neugierigen MLA, die noch zu interpretierende Eindrücke ermöglichen.

Abgrenzung technischer Unterstützung von Kunst und Wissenschaft

Nachdem überprüft wurde, ob die technische Unterstützung mittels MLA einen relevanten Beitrag bei der Welterzeugung leisten kann, besteht der letzte Schritt darin, diese Unterstützung innerhalb von Goodmans Perspektive zu verorten. Goodman selbst nennt drei stark verwandte Zweige der Welterzeugung: die Kunst, die Wissenschaft und die Wahrnehmung.

[...] my insistence on the very continuity and unity, the very affinity, of art and science and perception as branches of worldmaking.

(Goodman 1978, S. 133)

Die Diskussion maschinellen Lernens im Kontext von Goodmans Perspektive hat sich aufgrund der Schwerpunkte in dessen Argumentation vorwiegend am Zweig der Kunst orientiert, aber die technische Unterstützung stellt wie bereits erwähnt keine eigene, der Kunst ebenbürtige Möglichkeit dar, Welt zu erzeugen. Stattdessen *unterstützt* maschinelles Lernen im Rahmen einer typischen Anwendung – etwa im **Data Mining** – die Wahrnehmung des Nutzers. Gelegentlich werden MLA auch zur Unterstützung im wissenschaftlichen Kontext eingesetzt, während ein Einsatz durch Künstler (Brown et al. 2007) oder zur direkten Erstellung von Kunst (Aguilar et Lipson 2008) nur selten auftritt. Goodmans Ausspruch über die Verschränkung der Zweige der Welterzeugung „Painting is a science of which pictures are the experiment“ (Goodman 1978, S. 139) lässt sich entsprechend umformulieren in den Ausspruch, dass das Beobachten eine Wissenschaft ist, bei der Autoadaptionsprozesse die Experimente sind. Über diese Formulierung tauchen hier die Überlegungen zur Abgrenzung vom Begriff des **Experimentes** wieder auf und der Einsatz von MLA zu Ermöglichung von Experimenten bietet sich als ein möglicher Anknüpfungspunkt an diese Diskussion an.

3.6.3. Angemessenheit von Strukturvorschlägen

Der letzte Aspekt von Goodmans Perspektive, der im Kontext des maschinellen Lernens diskutiert werden soll, ist sein Umgang mit dem Konzept der Wahrheit. Goodman schlägt mit dem Begriff der Angemessenheit ein Bewertungskriterium vor, das nicht auf einen direkten Nutzen abzielt und auch für neugierig maschinell lernende Artefakte Verwendung finden könnte.

Der Wahrheitsgehalt eines Strukturvorschlages könnte als dessen Übereinstimmung mit den Rohdaten verstanden werden. Eine solche Begriffsbildung wäre jedoch nicht hilfreich, da solchermaßen wahre Strukturvorschläge sehr anfällig für eine Überanpassung wären. Eine derart bestimmte Wahrheit wäre zwar erstrebenswert, aber es wäre im Kontext des maschinellen Lernens kein Absolutheitsanspruch mitgedacht. Rohdaten basieren häufig auf Messwerten, die ein Rauschen enthalten. Eine 98%-Wahrheit, bei der der Strukturvorschlag in 98% der Fälle mit den Rohdaten übereinstimmt, wäre insbesondere in Kontexten dieser Art deutlich hilfreicher als eine 100%-Wahrheit. Die Ausdrucksfähigkeit von 98%-wahren Strukturvorschlägen wäre deutlich besser als diejenige von wahren Strukturvorschlägen und das entspricht nicht den Assoziationen, die der Begriff der Wahrheit wecken soll. Goodman führt als Alternative zu Wahrheit die Begriffe der **Richtigkeit** und der **Angemessenheit** ein. Im Rahmen der Diskussion und Suche des *Fair Sample* formuliert er seine Begriffsbildung zur Angemessenheit von Kunstwerken folgendermaßen.

In other words, rightness of design, color, harmonics – fairness of a work as a sample of such features – is tested by our success in discovering and applying what is exemplified. What counts as success in achieving accord depends upon what our habits, progressively modified in the face of new encounters and new proposals, adopt as projectible kinds. A Mondrian design is right if projectible to a pattern effective in seeing a world.

(Goodman 1978, S. 137)

Die Angemessenheit von Strukturvorschlägen basiert auf der Richtigkeit der Vorstruktur und äußert sich in der erfolgreichen Interpretation der Strukturvorschläge durch die Nutzer. Weiter kann neben der hier genannten progressiven Weise der Modifizierung der Wahrnehmung auch eine konservative Modifizierung gedacht werden, wenn keine neuen, sondern bereits bekannte Kontexte betrachtet werden (Weber 2010, S. 10). Im Falle einer konservativen Modifizierung der Wahrnehmung werden die im jeweiligen Kontext herrschenden Konventionen vom entstehenden Strukturvorschlag respektiert, wie es etwa in der zielorientierten Erstellung von Strukturvorschlägen mittels einer Stützvektormethode der Fall ist. Generell wird eine konservative Modifizierung eher mit Hilfe von mathematischer Optimierung und zielorientiertem Lernen realisiert werden können, da vor allem komplexe, mathematisch modellierte Kontexte die Einhaltung sehr vieler Konventionen erfordern und erlauben. Solche Konventionen können neugierigen MLA zwar mit Hilfe der Vorstruktur ebenfalls mitgegeben werden, allerdings wird ein neugieriges MLA seine Vorstruktur gegebenenfalls manipulieren oder an den entstandenen Strukturvorschlägen Manipulationen durchführen, die zwar den Rohdaten, nicht aber den Vorgaben der Vorstruktur entsprechen.

Insgesamt bietet die Rede von Angemessenheit und Richtigkeit eine Möglichkeit, den Grad der Wünschenswertheit einer Lernstrategie beschreiben und damit die Eignung für den

jeweiligen Kontext einschätzen zu können. Ein erwünschter Strukturvorschlag wäre ein angemessener Strukturvorschlag, da er sich interpretieren lässt. Nutzer von neugierigen MLA wünschen sich dementsprechend angemessene, unvorhergesehene Strukturvorschläge, während Nutzer zielorientierter MLA Wert auf die Richtigkeit der Vorstruktur legen. Eine weitergehende Ausarbeitung der Begriffe der Angemessenheit und Richtigkeit scheint vielversprechend und soll hier als Anknüpfungspunkt für weitere Projekte festgehalten werden.

3.7. Zusammenführung der Ergebnisse

Die wesentlichsten Ergebnisse der im Vorherigen durchgeführten Bestimmungen des zweiten Hauptteils lassen sich wie folgt zusammenfassen: Das erste Ergebnis ist die Beschreibung maschineller Lernvorgänge als Autoadaptionsprozesse, die sich als die Adaption einer Vorstruktur und eines Strukturvorschlags modellieren lassen. Das zweite Ergebnis liegt im verbesserten Auflösungsvermögen und den präziseren Beschreibungsmöglichkeiten, die sich aus der Unterscheidung von zielorientiertem und neugierigem Lernen ergeben. Das dritte Ergebnis ist die Charakterisierung der aus neugierigen Autoadaptionsprozessen gewonnenen Resultate als unvorhergesehene Strukturvorschläge, die von einem nutzerseitigen Wunsch nach Angemessenheit begleitet werden. Im Folgenden werden speziell die Unterscheidung zwischen neugierigem und zielorientiertem maschinellern Lernen und die präzisierten Beschreibungsmöglichkeiten rekapituliert und anschließend genutzt, um das Konzept der Welttechnik zu entwickeln und eine Möglichkeit des Zugangs zu Nichtwissen mittels MLA darzustellen.

Die sowohl bei neugierigem als auch bei zielorientiertem maschinellern Lernen zum Einsatz kommenden Autoadaptionsprozesse konnten im Vorhergehenden als ein iterativer Prozess beschrieben werden, der gemäß einer speziellen Vorstruktur abläuft und einen Strukturvorschlag erstellt, der Eingaben Ausgaben zuordnet. Die Vorstruktur umfasst und modelliert die Rahmenbedingungen wie etwa im Vorfeld bereits bekanntes theoretisches Hintergrundwissen. Diese Vorstrukturierungen können strikte Vorgaben darstellen und dem Autoadaptionsprozess einen speziell festgelegten oder standardisierten Ausgangszustand vorgeben. Alternativ können die Vorstrukturierungen selbst wiederum stark autoadaptiv gestaltet werden. Die Bedeutung der Vorstruktur ergibt sich daraus, dass ein maschinell lernendes Artefakt nach dem Ende des Autoadaptionsprozesses zwar auf Eingabereize auf eine beobachtbare und anhand von Testeingaben bewertbare Weise reagiert, dass jedoch eine Analyse, ob die Reaktionen auch in noch nicht getesteten Kontexten wunschgemäß ausfallen würden, schwierig oder nahezu unmöglich sein kann. In der Praxis ist es durchaus möglich, dass ein Artefakt ausschließlich bezüglich der bereits erhaltenen Eingabereize wunschgemäß reagiert (Dwyer 2005; Stallkamp et al. 2011). Das heißt, im Vorfeld einer Nutzung solcher Artefakte muss häufig analysiert werden, warum die Artefakte so reagieren, wie sie reagieren. Diese Analyse dient insbesondere dazu, ein Verständnis der jeweiligen Vorstrukturen zu schaffen. Darüber hinaus wurde in der Diskussion dargestellt, dass von MLA präsentierte Strukturen entweder schon bei der Festlegung der Vorstruktur als Repräsentationen gedacht werden, oder dass erst nach Abschluss des Autoadaptionsprozesses eine Interpretation der Strukturvorschläge durch den Menschen vorgenommen wird.

Zielorientierte Autoadaptionsprozesse wurden im Vorherigen als solche bestimmt, die vorformulierte Probleme lösen, indem sie klar definierte Ziele erreichen. Meist nähern sie sich ihren Zielen dabei entsprechend eines ebenfalls im Vorhinein definierten Qualitätsmaßstabes. Die Bewertung mittels dieses Maßstabes kann dabei Teil der Lernstrategie oder dem Autoadaptionsprozess nachgelagert sein. Ein Beispiel für eine auf ein Ziel ausgerichtete Problemlösung war die Suche nach möglichen Produktempfehlungen für Kunden auf Basis von deren bisherigem Einkaufsverhalten¹⁴¹. Solcherart vorstrukturierte und optimierende

¹⁴¹ Das Ziel ist hier sehr klar im monetären Gewinn als Qualitätsmaßstab messbar und die Gründe für das Kaufverhalten spielen für die Zielerreichung keinerlei Rolle.

maschinell lernende Artefakte konnten als transklassische Technik identifiziert und als zielorientiertes maschinelles Lernen beschrieben werden.

Die ehemals funktionsorientierte Technik wird, so die Forderung, zu einer zielorientierten Technik, die auf einer adaptiv gewonnenen Informationsbasis antizipatorisch die Problemlösungen vornimmt und dabei koordinierend/vernetzend die Problemlösungen Dritter in Rechnung stellt (»peer to peer«).

(Hubig 2007, S. 47)

Eine zweite Variante maschinellen Lernens konnte bei MLA beobachtet werden, deren Autoadaptionsvorgang neben einem systematischen auch einen signifikanten zufälligen Anteil enthielt. Diese zufälligen Anteile werden erst im Nachhinein als Konzepte interpretiert beziehungsweise wahrgenommen. In jedem Fall lassen sie eine Analyse des Artefaktes schwierig erscheinen und für Artefakte wie nicht vorstrukturierte KNN, die einen großen zufälligen Anteil aufweisen, ist eine Analyse nicht nur schwierig sondern fast ausgeschlossen. Solche Artefakte wurden als neugierig maschinell lernende Artefakte bezeichnet. Die Stärke und Besonderheit von neugierigen Autoadaptionsprozessen, die im Rahmen ihrer Vorstruktur nur wenige Vorgaben erhalten haben, liegt in der Möglichkeit, ein entsprechendes Artefakt unvorhergesehene und damit völlig unbekannte Strukturvorschläge suchen zu lassen. Derartige Artefakte ermöglichen nicht nur ein erweitertes technisches Handeln, sondern erschließen vollständig neue Handlungsumfelder, indem sie eine technische Form der Neugier realisieren. Die Hoffnung oder der Wunsch ist, dass der Nutzer sich in der Lage sieht, die entstandenen Strukturvorschläge aufzunehmen, dass er von ihnen kreativ angeregt wird und selbst spontan eine Möglichkeit sieht, wie er das Ergebnis in irgendeiner Art und Weise interpretieren kann. Eben gerade so, als würde die eigene Neugier zur Inspiration genutzt. Soll dies gelingen, benötigt der Nutzer theoretisches oder heuristisches Wissen, um das Resultat des Autoadaptionsprozesses überhaupt auf irgendeine Art und Weise weiterverarbeiten zu können. Allerdings kann hier nicht mehr von der Struktursuche als einem Zweck des Einsatzes eines maschinell lernenden Artefaktes gesprochen werden. Bei der skizzierten ziellosen Nutzung kann der Nutzungswunsch nicht die Struktur des Autoadaptionsprozesses selbst betreffen, sondern er bezieht sich auf die Möglichkeiten der Interpretation des Lernergebnisses. Entsprechend können in solchen Fällen bezüglich des Autoadaptionsprozesses beziehungsweise des Strukturvorschlages auch keine Soll-Ist-Vergleiche angestellt werden – das Resultat kommt in jeder Hinsicht unerwartet und unvorhergesehen. Wie bereits diskutiert wurde, sind Sachverhalte nur insofern Zwecke, als ihre Herbeiführbarkeit durch mögliche Mittel unterstellt werden kann (Hubig 2007, S. 231). Entsprechend kann bei einem neugierigen Einsatz maschinellen Lernens allenfalls noch von Wünschen gesprochen werden. Im Rahmen eines solchen zwecklosen Einsatzes von maschinell lernenden Artefakten kann ein Nutzer den Autoadaptionsprozess oder sein eigenes Verhalten entsprechend nicht auf Basis der Nutzungswünsche reflektieren. Neugierige MLA eignen sich aufgrund ihres zufallsähnlichen Vorgehens besonders für Einsatzgebiete, in denen eine Vorstrukturierung komplex oder nicht gewollt ist. Diese Situation tritt insbesondere bei der Arbeit mit riesigen Datenbanken unbekannter Struktur auf und ein anschauliches Beispiel für die Relevanz beider Varianten maschinellen Lernens findet sich in genau solch einer

Anwendung¹⁴². Im Rahmen des amerikanischen Präsidentschaftswahlkampfes 2012 wurde im politischen Kontext erstmalig in großem Maßstab Data Mining eingesetzt. Die Kampagne brachte systematisch und erfolgreich zielorientiertes Lernen zum Einsatz, um von Experten bereits vermutete Zusammenhänge zu verifizieren.

In late spring, the backroom number crunchers who powered Barack Obama's campaign to victory noticed that George Clooney had an almost gravitational tug on West Coast females ages 40 to 49. The women were far and away the single demographic group most likely to hand over cash, for a chance to dine in Hollywood with Clooney — and Obama.

So as they did with all the other data collected, stored and analyzed in the two-year drive for re-election, Obama's top campaign aides decided to put this insight to use. They sought out an East Coast celebrity who had similar appeal among the same demographic, aiming to replicate the millions of dollars produced by the Clooney context. "We were blessed with an overflowing menu of options, but we chose Sarah Jessica Parker," explains a senior campaign adviser. And so the next Dinner with Barack contest was born: a chance to eat at Parker's West Village brownstone.

(Scherer 2012, S. 1)

Gleichzeitig konnte im Rahmen von spontanen und kreativen Suchen in den Rohdaten auch sehr erfolgreich neugieriges maschinelles Lernen genutzt werden.

The numbers also led the campaign to escort their man down roads not usually taken in the late stages of a presidential campaign. In August, Obama decided to answer questions on the social news website Reddit, which many of the President's senior aides did not know about. "Why did we put Barack Obama on Reddit?" an official asked rhetorically. "Because a whole bunch of our turnout targets were on Reddit."

(Scherer 2012, S. 2)

Als weitere Veranschaulichung der Unterscheidung zwischen neugierigen und zielorientierten Artefakten diene in der Diskussion der optionalen Zwecklosigkeit der Struktursuche ein Schachprogramm, dessen Strukturvorschlag die Fertigkeit entwickeln soll, in jeder Stellung einen guten Zug empfehlen zu können. Falls das Entwicklungsziel darin besteht, ein kommerzielles Schachprogramm zu schreiben, wäre der Autoadaptionsprozess zielorientiert abgeschlossen. Allerdings könnte das Ziel auch sein, als Nutzer mehr über Schach zu lernen. In diesem Fall nützen die Zugvorschläge nichts, solange nicht ein Experte anschließend durch eine Interpretation ein Konzept entwickeln kann, das die Gründe offenlegt, aufgrund derer die empfohlenen Züge mit großer Wahrscheinlichkeit zum Sieg führen. Die Implementierung des Schachprogramms würde hier einmal zielorientiertem und einmal neugierigem maschinellem Lernen entsprechen. In den beiden Szenarien würden sehr unterschiedliche Lernstrategien zum Einsatz kommen, da auch sehr unterschiedliche Aspekte des entstehenden

¹⁴² Zielorientiertes maschinelles Lernen tritt in so gut wie jedem Kontext auf, in dem maschinelles Lernen eingesetzt wird, insbesondere auch bei der Analyse riesiger Datenbanken.

Schachprogramms von besonderer Bedeutung wären. Hieran ist gut zu erkennen, dass die Unterscheidung zwischen den beiden Varianten maschinellen Lernens nicht primär die Reaktionen des entstehenden Strukturvorschlages betrifft, sondern den Autoadaptionsprozess, über den diese Reaktionen erlernt wurden. Weiter gibt die Unterscheidung eine Auskunft darüber, wie das Lernresultat nutzbar ist, beziehungsweise wie ein Nutzer den Lernvorgang beeinflussen oder gar steuern könnte. In beiden Fällen ist in erster Linie das erlernte Verhalten und weniger der formale Zustand des Strukturvorschlages von Interesse, da dieser so gut wie sicher nicht intelligibel ist.

3.7.1. Zugang zu Nichtwissen

Sowohl zielorientierte als auch neugierige Artefakte unterscheiden sich auf den ersten Blick deutlich von dem früher beschriebenen Begriff klassischer Technik, weil keine **Steuer- und Regelungsprozesse** identifiziert oder gar unterschieden werden können. Zwar entspricht zielorientiertes maschinelles Lernen recht gut den vorgestellten Kriterien der transklassischen Technik, das Gleiche gilt jedoch nicht automatisch auch für neugieriges maschinelles Lernen. Die Unterscheidung zwischen klassischer und **transklassischer Technik** stellt keine vollständige Disjunktion dar. Zwischen den Begriffen besteht noch eine Form von Niemandsland und die Unterscheidung ist zunächst nur von der klassischen Technik aus gedacht. Nachdem sich neugierige Artefakte aus Perspektive der klassischen Technik als nicht-klassisch darstellten, verblieb somit die Frage, wie sich neugierige Artefakte aus Perspektive der transklassischen Technik verorten lassen. Dazu wurde eine positive Darstellung der transklassischen Technik als Raum von Technik erstellt. Die Ziellosigkeit neugieriger Artefakte und die Unvorhersehbarkeit der entstehenden Strukturvorschläge bedingten, dass bei einem einzelnen neugierigen Strukturvorschlag weder von klassischer noch von transklassischer Technik gesprochen werden kann. In der Abgrenzung von der **Informationstechnik** wurden isolierte Strukturvorschläge daher als eine Form von algorithmischem Gestalten beschrieben¹⁴³. Die Motivation zur Entwicklung von maschinell lernenden Artefakten liegt dementsprechend in deren sehr hoher Arbeitsgeschwindigkeit. Artefakte dieser Art sind in der Lage, in kurzer Zeit extrem viele mögliche Strukturvorschläge und damit mögliche Weltbezüge auf Basis einer formlosen und damit sonst nicht beobachtbaren Datenquelle zu erstellen. Weiter wurde diskutiert, dass der digitale Charakter von MLA es erlaubt, die Artefakte sehr effizient mit nachgelagerten automatischen Analysen zu kombinieren, die die Strukturvorschläge prüfen oder weiterverarbeiten. Die Nutzung solcher automatischen Analysen verbessert die aus der Arbeitsgeschwindigkeit der MLA resultierende Fähigkeit zur Arbeit mit einer Vielzahl von Strukturvorschlägen noch zusätzlich. Die Nutzung automatischer Analysemethoden lässt sogar den Eindruck entstehen, dass der Wunsch nach Angemessenheit des vormals isolierten Strukturvorschlages über die Verfügbarkeit heuristischer Erfolgswahrscheinlichkeiten zu einer Erwartung wird. Solch eine Rückgewinnung der Erwartbarkeit ist jedoch nicht in jedem Fall möglich, insbesondere, wenn die Erschließung vollständig unbekannter Rohdaten angestrebt wird, allerdings ist der resultierende Gedanke, dass neugieriges Lernen gegebenenfalls doch Kriterien klassischer Technik erfüllt, dennoch in gewisser Hinsicht belastbar. Ein weiteres Argument hierzu entsteht daraus, dass der Umgang mit neugierigen Artefakten ein Sich-in-ein-Verhältnis-Setzen zu den Ergebnissen des Autoadaptionsprozesses ermöglicht beziehungsweise erzwingt, anstatt es außer Kraft zu setzen. Dieses Außerkraftsetzen stellt

¹⁴³ Dies auch nur, wenn sie ein Mensch mit dem Strukturvorschlag in ein Verhältnis setzt.

eine der Motivationen zur Betrachtung transklassischer Technik dar, das heißt, eine Technik, die diese Eigenschaft nicht zeigt, weicht damit deutlich von der Idee hinter der transklassischen Technik ab – womit ebenfalls der Eindruck einer Nähe zur klassischen Technik entsteht. Beide Argumente deuten darauf hin, dass hier ein zentrales begriffliches Merkmal klassischer Technik in höherstufiger Form vorliegt. Neugieriges maschinelles Lernen ließ sich in der Diskussion transklassischer Technik entsprechend nicht nur nicht als eine solche verorten, sondern es erschien sogar wieder als eine Form klassischer Technik, die höherstufige Handlungsoptionen ermöglicht. Die hieraus motivierte These lautet, dass neugierige Artefakte eine höherstufige Form von Zugang zu unvorhergesehenen Strukturen darstellen und somit eine Möglichkeit des Umgangs mit **Nichtwissen** bieten (Harrach 2013). Neugierige MLA können dementsprechend eingesetzt werden, um Nichtwissensbereiche höherstufig zugänglich zu machen – nicht in dem Sinne, dass Nichtwissensbestände in Wissensbestände transformiert werden, sondern dass Nichtwissensbestände höherstufig strukturiert und zur Interpretation präsentiert werden – im einfachsten Fall mittels einer Gruppierung beziehungsweise der Erstellung von **Clustern** von Rohdaten in einem vormals ungreifbaren Datenmeer. Auch aufbauend auf dieser Perspektive ist es jedoch nicht möglich, neugierige MLA gezielt einzusetzen. Der Wunsch nach **Angemessenheit** entspricht dem Streben nach Zugang zu Nichtwissensbereichen. Wenn das Vorliegen eines unbekannten Bereichs bekannt oder vermutet wird, können neugierige MLA eingesetzt werden, um den Horizont der Nutzer im jeweiligen Kontext zu erweitern, allerdings kann in solchen Kontexten entweder nicht von Nichtwissensbereichen gesprochen werden oder die Strukturvorschläge sind unvorhergesehen und entsprechen damit gerade nicht den Vermutungen über die möglichen Anknüpfungspunkte an das bereits bekannte Vorwissen. Insgesamt können neugierige MLA zwar auch in Kontexten zum Einsatz kommen, in denen eine vergleichsweise klare Erwartung formuliert werden kann, wie der entstehende Strukturvorschlag aussehen wird, allerdings würde in der Praxis in diesen Fällen sehr wahrscheinlich ein zielorientierter Ansatz vorgezogen werden. Die Diskussionen der Rede von Nichtwissen ist damit natürlich nur angedeutet, kann allerdings als interessanter Anknüpfungspunkt an die Diskussion maschinellen Lernens festgehalten werden.

Eine Nebenbetrachtung der Analyse der **Erwartbarkeit** von Strukturvorschlägen besteht in der Frage, ob ein Suchvorgang und damit insbesondere eine technisch unterstützte Suche überhaupt damit enden kann, dass nichts gefunden wird. Dies ist insbesondere insofern interessant, als das Wissen, dass ein Ergebnis entstehen wird, die Intuition fördert, dass das Ergebnis in gewisser Weise erwartbar und nicht unvorhergesehen ist. Prinzipiell kann diesbezüglich argumentiert werden, dass etwa ein KNN erst dann seinen Autoadaptionsprozess beendet, wenn gewisse Stabilitätsforderungen der Vorstruktur erfüllt sind. Das heißt, es ist möglich dass spezielle Rohdaten einer Erreichung dieser Stabilitätsforderungen systematisch entgegenstehen und dass der Autoadaptionsprozess dementsprechend zumindest theoretisch niemals abgebrochen werden könnte. Ein Beispiel wäre die Übergabe rein zufälliger Rohdaten, die das MLA dazu veranlassen, seine Gewichte ständig zwischen verschiedenen Zuständen hin und her zu verschieben. Das KNN schlägt in diesem Fall in jedem Iterationsschritt eine gewisse Struktur vor und diese Struktur verändert sich mit jedem weiteren Iterationsschritt wieder. In der Praxis würde solch ein Autoadaptionsprozess vom Nutzer jedoch nach endlicher Zeit abgebrochen werden und der Nutzer würde ein Urteil fällen, sei es auch, dass die Rohdaten nur aus einem Rauschen bestehen oder dass das dynamische Verhalten des KNN als ein rhythmisches Muster interpretiert wird. Das Problem besteht nun darin, dass in jeder endlichen Signalfolge eine

Systematik gefunden werden kann¹⁴⁴. Zurückgespiegelt auf die Auswertung von **SETI** Daten¹⁴⁵, wäre durch den endlichen Betrachtungszeitraum immer eine Interpretation des Strukturvorschlags als außerirdische Nachricht möglich. Die Nutzer würden in solchen Situationen gegebenenfalls sehr lange glauben, sie hätten die Struktur der Signale nur noch nicht verstanden und hätten große Schwierigkeiten abschließend festzustellen, dass sie nichts gefunden haben. Dieser Sachverhalt lässt jedoch keine Rückschlüsse auf die Erwartbarkeit oder Unvorhersehbarkeit von Strukturvorschlägen zu. Wenn die Möglichkeit nichts zu finden offen gehalten werden soll, müssen „nur“ die Stabilitätsforderungen der Vorstruktur analysiert und respektiert werden.

3.7.2. Neugieriges maschinelles Lernen als Technik

Die Frage, ob es sich bei neugierigem maschinellem Lernen um Technik handelt, konnte durch die Betrachtungen zur Erzeugung von Welt in Abschnitt 3.6 vorsichtig bejaht werden, allerdings ist auch nach Abschluss dieser Betrachtung unklar, welche Art von Technik vorliegt. Der Begriff der Informationstechnik wurde bereits als nicht zielführend zurückgewiesen, da im Rahmen des Einsatzes neugieriger MLA selbst im angedeuteten heuristischen Kontext aus den Strukturvorschlägen erst durch menschliche Interpretationen Informationen werden. Weiterhin werden neugierigen Artefakten keine Informationen übergeben, sondern Rohdaten, die lediglich messbar sind, jedoch keine im jeweiligen Kontext erkennbare Struktur aufweisen. Auch auf grundlegenderer Ebene war eine Verortung maschinellen Lernens als Technik nicht erfolgreich. Insbesondere wurde etwa festgestellt, dass maschinelles Lernen insgesamt aufgrund der fehlenden Erwartbarkeit, Steuerbarkeit und Reflektierbarkeit des Autoadaptionsprozesses sowie der resultierenden Strukturvorschläge nicht als klassische Technik bezeichnet werden kann. Gleichwohl konnte dennoch nur zielorientiertes maschinelles Lernen als transklassische Technik verortet werden.

In Reaktion auf dieses Beschreibungsdefizit wird an dieser Stelle, aufbauend auf der Betrachtung der Erzeugung von Welt in Abschnitt 3.6, das Auftreten einer neuartigen Form von Technik konstatiert, die durch die Hochgeschwindigkeitserzeugung von aus Messwerten gewonnenen Strukturvorschlägen möglich wird und die eine technische Unterstützung bei der Welterzeugung anbietet. Diese Form von Technik wird als **Welttechnik** bezeichnet¹⁴⁶. Welttechnik zeichnet sich insbesondere dadurch aus, dass sie dem Nutzer einen ungezielten Zugang zu Bereichen des Nichtwissens eröffnen kann. Artefakte der Welttechnik erzeugen Strukturvorschläge, die Möglichkeiten darstellen, Rohdaten zu strukturieren und damit überhaupt erst wahrzunehmen. Welttechnik dient somit der Herstellung von Optionen von Weltbezüglichkeit. Gerade neugierig maschinell lernende Artefakte stellen somit vielfach Werkzeuge zur Welterzeugung dar (Goodman 1978). Die Artefakte können daher statt als

¹⁴⁴ Dies kann mit Hilfe eines Gedankenexperiments veranschaulicht werden, bei dem die kleinste uninteressante Zahl gesucht wird. Jede Zahl, die diese Eigenschaft hätte, wäre genau dadurch, dass sie diese Eigenschaft aufweist – aus mathematischer Sicht – wieder interessant.

¹⁴⁵ “SETI is an acronym for Search for Extraterrestrial Intelligence. It is an effort to detect evidence of technological civilizations that may exist elsewhere in the universe, particularly in our galaxy.” (SETI Institute 2013)

¹⁴⁶ Neugierige Artefakte könnten auch unter dem Begriff der neugierigen Technik zusammengefasst werden, die Bezeichnung als Welttechnik wird hier jedoch bevorzugt. Der Grund ist, dass Neugier angesichts der nützlichen Assoziationen und Intuitionen als Problembezug eingeführt wurde und die Unterstützung des Menschen bei Erzeugung von Weltbezug im Sinne Goodmans zunächst eine stabilere Basis darstellt. Der Begriff der Neugier soll damit keineswegs abgetan werden, es besteht hier lediglich noch ein gewisser Forschungsbedarf, bevor neugierige Technik als Bezeichnung mit dem Begriff Welttechnik konkurrieren kann.

Informationstechnik präziser als Welttechnik bezeichnet und verstanden werden. Welttechnik zeichnet sich insbesondere dadurch aus, dass sie keine Problemlösungen erstellen kann oder soll, sondern sich ausschließlich an Eingabereize anpasst und das Resultat dieser Anpassung dem Nutzer präsentiert. Somit reagiert Welttechnik nicht auf bekannte Probleme und funktioniert nicht auf Basis von Informationen, sondern unterstützt den Nutzer dabei, Weltbezüge zu entwickeln und Probleme erst als solche artikulieren zu können.

Das Konzept von Welttechnik ist nicht exklusiv an die Betrachtung von maschinell lernenden Artefakten gebunden. Die Frage, welche Aspekte des Konzeptes von Welttechnik spezifisch für MLA sind und inwieweit sich bestimmte Aspekte verallgemeinern lassen, ist noch offen und bildet einen Anknüpfungspunkt für weitere Analysen. Die Suche nach einer Übertragbarkeit der Idee von Welttechnik auf Kontexte außerhalb der Informatik könnte an die verschiedenen Formen von Selbstorganisation und Autoadaption anknüpfen, die auch anderen Gebieten technischer Artefakte zugeschrieben werden. Gerade eine Übertragung auf Biotechnik und mit etwas mehr Aufwand auf die Nanotechnologie und die Neurowissenschaften erscheint vielversprechend. Generell bieten sich Technikbereiche, die der transklassischen Technik zugeordnet werden, als Suchräume für weitere Formen von Welttechnik an. Eine solche Suche wäre für den jeweiligen Technikbereich auch dann ein Gewinn, wenn keine Welttechnik gefunden werden kann, da sich die jeweilige Technik in diesem Fall voraussichtlich als zielorientiert erweisen würde, was wiederum die Rede von Selbstorganisation präzisiert. Eine Suche nach Welttechnik kann entsprechend dazu beitragen zu analysieren, was die genannten und häufig als verwandt betrachteten Bereiche *tatsächlich* gemeinsam haben. Der prognostizierte Mehrwert der Suche basiert auf der These, dass ein Bereich, der neugierige und zielorientierte Technik enthält, nicht undifferenziert diskutiert oder gar charakterisiert werden kann. Dieser These folgend wäre es überaus erstrebenswert, die Welttechnik gezielt aus der Diskussion der konvergierenden Technik zu exkludieren, da die Herausforderungen und Möglichkeiten der Diskussion von Welttechnik sehr speziell sind und gerade nicht mit denjenigen von transklassischer Technik übereinstimmen. Welttechnik stellt somit einen konkreten Beitrag für die Präzisierung der vagen und in vielen Fällen unscharfen oder gar widersprüchlichen Rede von „Selbstorganisation“ dar. Die These lautet, dass Diskussionen über „Selbstorganisation“ davon profitieren würden, wenn analog zu dieser Analyse des maschinellen Lernens als Vertretung der Informatik auch für die anderen Teilgebiete der **NBIC** – die Nanotechnik, Biotechnik und Kognitionswissenschaften – eine interdisziplinäre Grundlagenforschung ausgearbeitet würde, die gezielt beleuchtet, welche technischen Ansätze und Artefakte von Interesse sein könnten und welche Aspekte aus der Diskussion ausgeschlossen werden sollten.

3.7.3. Welttechnik in der Praxis

Aufbauend auf den zusammengeführten Ergebnissen wird im Folgenden ein Beispiel für ein Forschungsprojekt betrachtet, bei dem maschinelles Lernen in großem Umfang zum Einsatz gekommen ist. Das Ziel wird dabei sein darzustellen, wie über solch ein Forschungsprojekt nachgedacht werden kann und wie ein techniknaher Einstieg in eine Diskussion der technikphilosophischen Aspekte des Projektes gelingen kann. Das bedeutet, es werden im Folgenden weniger die Forschungsergebnisse des Projektes paraphrasiert, als dass diejenigen Punkte betont werden, die aus technikphilosophischer Sicht von Interesse sein könnten und die mit Hilfe der vorliegenden Aufarbeitung interdisziplinär greifbar geworden sind. Das betrachtete Beispiel (Zheng et al. 2009) versucht die Frage zu beantworten, was eine

Sehenswürdigkeit ist und wo auf der Welt sich welche Sehenswürdigkeiten befinden. Dieses Forschungsprojekt ist interessant, da die Begriffsbestimmung vergleichsweise ergebnisoffen unternommen wird und eine Vielzahl von Methoden des maschinellen Lernens zum Einsatz kommt.

Das Ziel des Forschungsprojektes liegt darin, eine umfassende Datenbank aller Sehenswürdigkeiten weltweit zu erstellen und dafür ein Artefakt zu konstruieren, das in der Lage ist, Fotos von Sehenswürdigkeiten als solche zu erkennen und die Datenbank zu editieren. Diese Zielstellung ist im Weiteren nur insofern wichtig, als sie das Verständnis für die Rahmenbedingungen des Forschungsprojektes verbessert. Der Fokus der Analyse des eingesetzten maschinellen Lernens liegt in der später dargestellten Hypothesenprüfung. Zu diesem Zweck müssen jedoch zunächst kurz die Rahmenbedingungen und Vorgaben des Projektteams diskutiert werden.

Die erste Vorgabe besteht darin, dass das Projektteam zwar offen lässt, welche Teile der Welt als Sehenswürdigkeit zu bezeichnen sind, jedoch eine allgemeine Bestimmung des Begriffs Sehenswürdigkeit vorgeben muss, um keine völlig beliebigen Ergebnisse zu erhalten.

The touristic landmarks are easily recognizable and wellknown sites and buildings, such as a monument, church, etc.[...]. They are the pivotal part of people' tours, due to their notable physical, cultural and historical features.

(Zheng et al. 2009, S. 1)

Dies stellt eine vergleichsweise klare Vorstellung dessen dar, was unter einer touristischen Sehenswürdigkeit verstanden werden soll. Weiter wird vom Projektteam festgestellt beziehungsweise gesetzt, dass etwas Sehenswürdiges häufig fotografiert wird. Dies kann insofern als eine Folgerung aus der Definition betrachtet werden, als der Begriff der Sehenswürdigkeit nur im Kontext einer touristischen Reise definiert wurde. Ein resultierender systematischer Fehler besteht darin, dass nur solche Sehenswürdigkeiten erfasst werden, die von Touristen besucht werden, die Fotoapparate besitzen und nutzen. Dieser systematische Fehler wird vom Forschungsteam zumindest implizit thematisiert und spielt für die zu prüfende These keine große Rolle – wie im Folgenden noch begründet wird.

Die Beobachtung, die das Forschungsprojekt angestoßen hat, ist ebenfalls zentral für die Erstellung der Vorstruktur der einzusetzenden MLA und besteht im Folgenden.

The explosion of personal digital photography, together with Internet, has led to the phenomenal growth of landmark photo sharing in many websites like Picasa Web Album (picasa.google.com).

(Zheng et al. 2009, S. 1)

Diese Beobachtung führte zu dem folgenden, schon beschriebenen Ziel.

With the vast amount of landmark images in the Internet, the time has come for computer vision to think about landmarks globally, namely to build a landmark recognition engine, on the scale of the entire globe. This engine is not only to visually recognize the presence of certain landmarks in an image, but also contributes to a worldwide landmark database that organizes and indexes landmarks, in terms of geographical locations, popularities, cultural values and social functions, etc.

(Zheng et al. 2009, S. 1)

Interessant für diese Diskussion des maschinellen Lernens sind zwei Aspekte des Forschungsprojektes: zum einen die Hypothese, die auf der Ausgangsthese und dem formulierten Ziel basiert und zum anderen die Methode, die vom Projektteam gewählt wurde, um die Hypothese zu prüfen und im Erfolgsfall das formulierte Ziel zu erreichen. Die Darstellung dieser beiden Aspekte beginnt, wie die Darstellung des Forschungsprojektes, mit der Betrachtung der Methode, um die Hypothese nachträglich als implizit enthalten zu identifizieren. Diese strukturelle Analogie soll das Verständnis des Aufbaus des Projektes verbessern¹⁴⁷.

To achieve this goal, we explore two sources on the Internet: (1) the geographically calibrated images in photo sharing websites like picasa.google.com and panoramio.com; and (2) travel guide articles from websites, such as wikitravel.com.

(Zheng et al. 2009, S. 2)

Die Analyse von Reiseberichten und Reiseführern stellt eine zielorientierte Suche in einem kontextreichen und stark verrauschten Bereich dar, die einiges Vorwissen voraussetzt, das den entsprechenden MLA über die Gestaltung der Vorstruktur mitgegeben werden muss. Die parallel durchgeführte Auswertung von Fotografien stellte vermutlich zumindest zu Beginn des Forschungsprojektes eine neugierige Suche dar, bei der Fotodatenbanken an MLA übergeben wurden und vergleichsweise ergebnisoffen auf Strukturvorschläge gehofft wurde, die eine angemessene Beschreibung von so etwas wie einer Sehenswürdigkeit darstellen. Zwar wird im Projektbericht der strategische Einsatz einer Reihe stark vorkonzeptionierter Autoadaptionsprozesse vorgestellt, allerdings ist stark zu vermuten, dass bei der Konzeptionierung der Strategie zunächst deutlich experimenteller vorgegangen wurde. Das Projektteam konnte wahrscheinlich konkrete Aspekte seiner Gesamtstrategie erst spät formulieren – etwa die explizite Modellierung der Ortsbestimmung einer Sehenswürdigkeit als in einer Stadt befindlich, die in einem Land liegt, das wiederum Teil eines Kontinents ist¹⁴⁸. Weiter war vermutlich erst nach der Ausarbeitung der Gesamtstrategie eine Optimierung der Autoadaptionsprozesse möglich, die in Konsequenz zu einigen der in der Veröffentlichung genannten Vorstrukturierungen führte. Dies stellt zwar prinzipiell nur eine

¹⁴⁷ Der Fokus auf die Darstellung der einzusetzenden Methodik gegenüber der Klärung der impliziten Vorannahmen stellt eine in Beiträgen der Informatik verbreitete Vorgehensweise dar, da häufig in erster Linie die eingesetzten Methoden von Interesse sind und die inhaltliche Anwendung als sekundär betrachtet wird.

¹⁴⁸ Die vom Projektteam gewählte Modellierung ist zwar nicht überraschend, aber es gibt sehr viele alternative Möglichkeiten der Darstellung.

Vermutung dar, allerdings wird der experimentelle Charakter der Vorgehensweise in der Formulierung der Hypothese explizit angesprochen und ein relativ ergebnisoffenes Arbeiten ist auch schlicht gute Forschungspraxis. Wenn man diese Vermutungen akzeptiert, lässt sich der Forschungsprozess in eine neugierige Startphase und eher zielorientierte Optimierungsphase aufteilen. Der Begriff der Sehenswürdigkeit wird zunächst allgemein formuliert und mittels neugieriger Strukturvorschläge konkretisiert, anschließend wird er vom Projektteam gezielt mit den Ergebnissen der Analyse der Reiseführer verglichen. Es liegt ein hochgradig reflektierter Umgang mit maschinellem Lernen vor, MLA werden zunächst ergebnisoffen in weitgehend unbekannten Kontexten eingesetzt und je mehr auf diese Weise mit den Rohdaten gearbeitet wurde, desto besser werden diese verstanden und die ganze Forschung wird zielorientierter. Kurz gesagt stellt dieses Forschungsprojekt eine auch interdisziplinär verständliche Analyse dar, die wahrscheinlich mit einem leicht vorstrukturierten, neugierigen Ansatz begonnen und im Anschluss zielorientiert optimiert wurde.

Die vorausgegangene Beschreibung der Arbeit des Projektteams harmoniert mit den Darstellungen maschinellen Lernens und ist auf viele praktische Beispiele von neugierigem Lernen übertragbar, die Motivation, gerade dieses Forschungsprojekt als Beispiel zu betrachten, liegt jedoch in der expliziten Nennung der Hypothese, die der Methode zugrunde liegt.

By exploiting these two sources of information, we can mine a more comprehensive list of landmarks. This is so because landmark is a perceptual and cognitive concept, which people of different background tend to perceive differently. Our experiments confirm this premise, by showing that the landmarks mined from GPS-tagged photos and travel guide articles have small overlap and complement each other.

(Zheng et al. 2009, S. 2)

Diese Formulierung gibt keine nachträglich erzeugte Rationalisierung wieder, sondern zeigt transparent den ergebnisoffenen Charakter der Suche und damit die Relevanz des Forschungsprojektes für die vorliegende Analyse des maschinellen Lernens. Das Projektteam sieht die eigene Vorgehensweise als Experiment, das dazu dient einen Begriff zu schärfen. Die Rede von einem Experiment anstelle eines Wunsches der Angemessenheit ist nachvollziehbar, aber nur genau dann korrekt, wenn im Rahmen der Forschungsarbeit keine irrelevanten „Experimente“ durchgeführt wurden¹⁴⁹ und die Aussage, dass die Reiseführer und die Fotodatenbanken signifikante und systematische Unterschiede aufweisen, bereits vor der Untersuchung antizipiert wurde. Unabhängig davon, ob die Rede von Experimenten an dieser Stelle angemessen ist, wurde zumindest rückblickend die These auf den Prüfstand gestellt und widerlegt, dass dasjenige, was große Gruppen von Menschen als Sehenswürdig betrachten, auch dasjenige ist, was von Reiseführern als Sehenswürdigkeit bezeichnet wird. Die Methode zur Beantwortung dieser Frage ist darüber hinaus *konstruktiv*, so dass ein Vergleich der unterschiedlichen Konzepte der Bestimmung von Sehenswürdigkeiten erstellt werden kann. Im Hinblick auf diese Hypothese wird auch klar, dass der beleuchtete systematische Fehler vernachlässigt werden kann, da die Nutzer der betrachteten englischsprachigen Webseiten

¹⁴⁹ Das inkludiert etwa auch, dass keine anderen Projekte zum gleichen Thema gescheitert sind.

sich zum allergrößten Teil aus Touristen zusammensetzen, die einen Fotoapparat oder ein entsprechendes Handy besitzen und das Mitführen und den Einsatz eines solchen Artefaktes auf einer touristischen Reise als normal betrachten.

Zusammenfassend lässt sich sagen, dass der Einsatz maschinellen Lernens eine Arbeit mit den riesigen Fotodatenbanken überhaupt erst möglich macht. Erst dieser Zugang ermöglicht den Forschern ein Verständnis der Rohdaten und befähigt sie dazu, hilfreiche Kategorisierungen vorzunehmen. Anders formuliert lässt der Einsatz maschinellen Lernens den im Projektbericht beschriebenen Gedankengang technisch überhaupt erst entstehen. Insgesamt ist dennoch auch bei diesem Forschungsprojekt mittels einer nachträglichen Außenansicht nicht mehr einwandfrei feststellbar, ob und inwiefern Welttechnik zur Unterstützung bei der Erstellung von Weltbezügen und zum Umgang mit Nichtwissen eingesetzt wurde. Wichtig ist jedoch, dass die Fragepfade, die zu diesem Zweck besritten werden müssten, offen liegen und somit ein interdisziplinärer Dialog mit der Forschergruppe möglich geworden ist.

3.7.4. Ununterscheidbarkeit der entstehenden Weltbezüge

Im vorangegangenen Beispiel wurde festgehalten, dass im Rahmen einer nachträglichen Betrachtung nicht feststellbar ist, ob die Erstellung der Weltbezüge der Forscher durch die Nutzung von Welttechnik unterstützt worden ist. Der Forschungsbericht des Projektteams konstatiert, mit Ausnahme der expliziten Nennung der Hypothese, lediglich eine Neuentdeckung und identifiziert beziehungsweise beschreibt somit genau wie beabsichtigt die Forschungsleistung.

Diese Wahrnehmung lenkt den Blick auf eine Frage, die im Rahmen der Betrachtung der Erzeugung von Welt in Abschnitt 3.6 bereits diskutiert wurde und die in dieser letzten Betrachtung des zweiten Hauptteils noch einmal explizit beleuchtet werden soll. Diese Frage war, ob Weltbezüge, bei deren Erstellung Welttechnik unterstützt hat, sich systematisch von anderen Weltbezügen unterscheiden und woran dieser Unterschied festgemacht werden könnte. Die These dazu war und bleibt, dass sich die Weltbezüge nicht unterscheiden und diese These soll im Weiteren anhand zweier Beispiele plausibilisiert werden. Das erste Beispiel bildet die Etablierung von Schachprogrammen, die zumindest aus der Perspektive von Großmeistern das menschliche Verständnis und die Art und Weise Schach zu spielen verändert haben.

I don't know whether computers are improving the style of play, I know they are changing it. Chess has become a different game, one could say that computers have changed the world of chess. That is pretty clear.

(Friedel im Interview mit Kramnik am 13.03.2006)

Die Spielweise dieser Schachprogramme wurde zunächst als unmenschlich wahrgenommen, insbesondere weil spezielle Anti-Schachprogramm-Spielweisen sehr erfolgreich waren. Eine einfache Richtlinie war und ist, dass das Spielen von eher ungewöhnlichen Zügen, die keinen unmittelbaren Nachteil bedeuten, sehr hilfreich ist.

Horrible as White's Pawn structure may be, Black can't exploit it because he'll be unable to develop his King's-side normally. It's the little quirks like this that could make life difficult for a chess machine.

(Fischer 1960)

Heute ist diese Ansicht weniger verbreitet, da die Programme ihre anfänglichen Schwächen abgelegt haben und ihre Spielweise von den Großmeistern analysiert und in ihr Spiel integriert wurde.

My match with Deep Junior is the beginning of a new era in computer chess. Both sides were under strict supervision and every aspect of each game was recorded. As a result, we can see the strengths and shortcomings of the machines, and of ourselves. This marks an important shift in the history of computer chess.

(Kasparov 2003)

Unabhängig davon spielen einzelne Schachprogramme zwar tatsächlich anders als einzelne Großmeister, aber auch die Spielweisen von Großmeistern unterscheiden sich, daher ist die bloße Abweichung kein ausreichendes Indiz für einen systematisch unmenschlichen Aspekt in der Spielweise.

What makes this new era so exciting is that there are many programs using different techniques that produce distinct styles. Deep Junior is as different from Deep Fritz as Kasparov is from Karpov.

(Kasparov 2003)

Auch heute ist das Spielen von unkonventionellen Zügen noch eine gute Vorgehensweise, um einen Vorteil im Spiel gegen ein Schachprogramm zu erlangen, allerdings ist diese Vorgehensweise nicht nur gegen Schachprogramme von Nutzen, sondern stellte auch schon vor dem Auftreten schachspielender Artefakte einen wichtigen Teil des strategischen Repertoires von Schachspielern dar. Insbesondere Garry Kasparov hat in diesem Bereich Stärken, während sich etwa Bobby Fischer mitunter nur schwer von seinem Stil lösen konnte, was gegen die aggressive Spielweise mancher seiner Gegner mitunter von Nachteil war. Dieses Aufeinandertreffen von Spielstilen macht auf einem bestimmten Level auch die Eleganz des Schachspiels aus, da es in solchen Situationen nicht mehr nur darum geht, Züge gegeneinander aufzurechnen, sondern auch darum, bestimmte Strategien, Strategiekomplexe oder Musterkomplexe zu entwickeln. Gerade die Weiterentwicklung dieses Aspekts des Schachspiels kann von Schachprogrammen unterstützt werden.

[The theory of the 1960s and 1970s] was not the same! Yes, we had a theory, which was well developed, but not the same.

[...]

The computer is important on the high level when the value of the move is high. It's very important to learn to work with the computer.

(Spassky 2009)

Es soll in diesem Zusammenhang kurz festgehalten werden, dass das Denken in abstrakten Strukturen und in Faktoren wie „Eleganz“ im Wettstreit mit MLA, die ein Brettspiel „erlernt“ haben, durchaus einen sinnvollen Ansatz darstellen kann. Noch konnten etwa keine **Go-Programme** entwickelt werden, die eine Siegchance gegen einen sehr guten menschlichen Spieler haben¹⁵⁰. Zwar ist es keineswegs ausgeschlossen, dass in einigen Jahren die ersten wirklich spielstarken Go-Programme entstehen, aber dennoch ist Go ein sehr eindrucksvolles Beispiel für ein Spiel, dessen Strategien auf abstrakten Analysen und intuitiven Eleganzkonzepten basieren. Wie im ersten Hauptteil dargestellt wurde, sind Algorithmen auch in Form von maschinellem Lernen nicht in der Lage, analytische Konzepte zu formulieren, sondern sind stattdessen gezwungen, diese numerisch zu approximieren. Die extrem große Anzahl möglicher Züge im Go bedingt dementsprechend, dass der Versuch, die Spielstärke eines Artefaktes allein über eine Steigerung der Rechengeschwindigkeit zu erhöhen, in absehbarer Zeit nicht erfolgreich sein wird.

Unabhängig von den Spezifika von Schach oder anderen Kontexten, wurde festgestellt, dass menschliche Spieler etwas daraus lernen konnten, dass sie sich zu der Spielweise von Schachprogrammen in ein Verhältnis gesetzt haben. Die Feststellung motiviert mitunter die Argumentation, dass eine Bedingung der Möglichkeit, dass solch ein Lernfortschritt auf Seiten der Spieler denkbar ist, darin besteht, dass ein systematischer Unterschied zwischen menschlichen Spielern und Schachprogrammen vorliegt, aus dem gelernt werden kann. Ein systematischer Unterschied der beiden Spielweisen hätte in dieser Argumentation eine Veränderung des Schachspiels zur Folge und das Auftreten einer Veränderung wiederum würde das Vorliegen des Unterschieds nachweisen. Diese Argumentation zerfällt schon auf den zweiten Blick, da das Schachspiel sich zwar verändert hat, die Veränderung aber in gewisser Hinsicht kontingent ist.

There is much more tactics involved, much more complications. These days, thanks to computers, in order to get anything out of the opening you have to go for complicated positions. Especially if you want to win the game. You cannot do so by simply getting a slightly better position and slowly grinding down your opponent. Well, actually you can, but it is becoming more and more difficult. So people are looking for complications.

(Friedel im Interview mit Kramnik am 13.03.2006)

¹⁵⁰ Go ist ein in Japan und Korea geprägtes 2-Personen-Brettspiel chinesischen Ursprung mit sehr einfachen Regeln und mit (auch und insbesondere im Vergleich zu Schach) einer sehr großen Anzahl möglicher Züge. Details finden sich in der Wikipedia (Wikipedia Contributors 2012, Go (Spiel)).

Zwar lag nicht irgendein beliebiger Trend vor, sondern der Trend entstand mit der Analyse der Vorgehensweise der Schachprogramme, aber es wäre ein anderer Trend entstanden, wenn moderne Großmeister andere Schwächen bei sich und ihren Gegnern entdeckt hätten. Es hätte sich etwa herausstellen können, dass Großmeister immer schon versucht haben die Komplexität der Stellungen zu erhöhen, um so ihre Fähigkeiten besser zur Geltung bringen zu können¹⁵¹ oder die Programme hätten Schach „lösen“ können¹⁵². Kurz gesagt hängen die denkbaren Veränderungen oder Konsequenzen aus der Analyse der Spielweise von Schachprogrammen sehr stark vom Hintergrund des interpretierenden Menschen ab.

Ein etwas anders gelagerter Versuch, das spezifisch Technische an technisch unterstützt entstandenen Weltbezügen zu bestimmen, argumentiert über die Möglichkeiten zur Manipulation der Nutzer. Als Veranschaulichung der entsprechenden Argumentation wird die Webseite Last.fm dienen. Nachdem vom Nutzer einige Musiktitel abgespielt wurden, die dieser besonders schätzt, erstellt der auf Last.fm hinterlegte Strukturvorschlag ein Radioprogramm aus Titeln, die dem Nutzer ebenfalls gefallen könnten. Diese Empfehlungsstruktur basiert auf einer umfangreichen Datenbank, die insbesondere aus den Rückmeldungen von registrierten Nutzern besteht, die bereits Titel bewertet haben. Die Argumentation der Suche nach einem typisch technischen Anteil von Last.fm könnte die These formulieren, dass der Musikgeschmack der Nutzer sich durch die Nutzung des Dienstes verändert. Diese These scheint plausibel und soll unter Vorbehalt übernommen werden. Analog zur Argumentation bei Schachprogrammen stellt eine solche Veränderung des Musikgeschmacks der Nutzer an sich noch keine relevante Beobachtung dar. Die Möglichkeit einer technischen Manipulation ergibt sich, wenn in sehr vielen Situationen Strukturvorschläge an den Nutzer herangetragen werden und der Nutzer sich dessen nicht oder zumindest nicht in Gänze bewusst ist. Möglicherweise wäre etwa nachweisbar, dass sich der Musikgeschmack der einzelnen Nutzer den Strukturen der Datenbank annähert¹⁵³. Im Bisherigen wurde jedoch der Umgang mit MLA und mit Welttechnik zumindest implizit als ein bewusster Umgang mit dem Artefakt beschrieben. Der Einsatz von Welttechnik und neugierigen Artefakten setzt gerade voraus, dass eine Umgebung als unüberblickbar empfunden wird und die Strukturvorschläge des Artefaktes vor diesem Hintergrund betrachtet werden. Wenn das Artefakt als solches dem Nutzer nicht gewahr ist, ist ein solcher neugieriger Einsatz nicht möglich. Das Sich-in-ein-Verhältnis-Setzen zu den Strukturvorschlägen ist ein zumindest impliziter Teil des Einsatzes von Welttechnik. Ein verdecktes Auftreten von Welttechnik ist möglicherweise nicht undenkbar, allerdings spielt es in der vorliegenden Diskussion keine Rolle. Der hier betrachtete Diskussionsrahmen würde dementsprechend verlassen, wenn die Möglichkeit einer technischen Manipulation des Nutzers analysiert wird. Die in diesem Fall zu betrachtenden Fragestellungen werden im Diskurs über das Ubiquitous Computing aufgearbeitet. Unabhängig davon, ob eine Form von neugierigem Umgang mit nicht wahrgenommenen Artefakten möglich ist, würde der Fokus im Falle der Analyse einer technischen Manipulation darauf liegen, dass der Nutzer sich nicht

¹⁵¹ Allerdings war und ist das Auftreten einer solchen Übereinstimmung extrem unwahrscheinlich.

¹⁵² Das heißt, sie hätten eine Siegstrategie für Weiß oder alternativ eine Remisstrategie für Schwarz finden können.

¹⁵³ Hier würde sich die Frage stellen, ob diese Beobachtung für viele oder gar alle denkbaren Datenbankstrukturen reproduzierbar wäre oder ob es ein Spektrum an möglichen Formen von menschlichen Musikgeschmäckern gibt und sich der einzelne Nutzer einer der Strukturen annähert, wenn er mit neuer Musik konfrontiert wird. Eine kurze Einführung in die Vorgehensweise von Empfehlungsalgorithmen und die Einbeziehung von Nutzerstereotypen bietet Weimer (Weimer et al. 2008, S. 2).

mehr bewusst mit dem Strukturvorschlag auseinander setzen kann und es würde folglich kaum noch eine Rolle spielen, wie der jeweilige Strukturvorschlag entstanden ist – sei es zielorientiert oder neugierig oder sei es durch einen Mitmenschen oder ein MLA. Zusammenfassend kann festgehalten werden, dass eine Argumentation mit Bezug auf die mögliche technische Manipulation des Nutzers durch maschinell lernende oder andere Artefakte unabhängig von den durch sie eröffneten Themenfeldern implizit eine Ununterscheidbarkeit von Weltbezügen, die mit und ohne Welttechnik entstanden sind, als gegeben annimmt.

Insgesamt scheint der Akt, die aus der Betrachtung eines Strukturvorschlages resultierende Veränderung im eigenen Denken dem Artefakt zuzuschreiben, wahlweise eine Romantisierung oder eine katastrophische Vorstellung des Nutzers darzustellen. Die aus der Interpretation eines Strukturvorschlages resultierende Veränderung im Denken dem den Strukturvorschlag erzeugenden Artefakt zuzuschreiben, würde ausblenden, dass ein reflektierter Umgang mit unvorhergesehenen Eindrücken durchaus erstrebenswert und nichts Außergewöhnliches ist. Es wäre etwa ein kurzschlüssig hergestellter Zusammenhang zu vermuten, dass die Webseite Last.fm bestimmte Veränderungen des Musikgeschmacks des Nutzers verursacht, nur weil sie ihren Strukturvorschlag basierend auf maschinell Lernen entworfen hat. Wenn der Nutzer die Webseite erwartungslos einsetzt, kann sie eine schwache Form von Welttechnik darstellen und den Nutzer mit für ihn unvorhergesehenen Bereichen von Musik konfrontieren. Eine solche Konfrontation entspräche der Zugänglichmachung eines besonders interessanten, unvorhergehörten Gebietes von Nichtwissen.

4. Ausblick

Ein Anliegen dieser Analyse des maschinellen Lernens besteht darin zu prüfen, welche Beiträge die Informatik und die Technikphilosophie für eine interdisziplinäre Diskussion des maschinellen Lernens bieten können. Zu diesem Zweck wurden wiederholt Anknüpfungspunkte und möglicherweise verwandte Technikbereiche aufgezeigt um zu identifizieren, was die nächsten Stationen einer Analyse sein können.

Einige dieser offenen Fragen und Anknüpfungspunkte werden im Folgenden etwas detaillierter ausformuliert.

4.1. Verallgemeinerbarkeit der Suche nach Welttechnik

Eine bereits mehrfach aufgetretene Frage war diejenige nach der Verallgemeinerbarkeit der Beobachtungen und Beschreibungen des zweiten Hauptteils auf andere Technikbereiche. Zwar basiert etwa speziell die im zweiten Hauptteil entwickelte Vorstellung einer Technik, die den Nutzer bei der Erstellung von Weltbezügen unterstützt, auf der interdisziplinären Aufarbeitung und Darstellung des maschinellen Lernens im ersten Hauptteil, dennoch wurde die so gefasste Welttechnik nicht auf die Informatik begrenzt gedacht. In diesem Zusammenhang entsteht die Frage, welche Aspekte von Welttechnik spezifisch sind für den technischen Kontext, aus dem sie gewonnen wurden, und inwieweit bestimmte Aspekte verallgemeinerbar sind – etwa für neugierige MLA wie nicht vorstrukturierte künstliche neuronale Netze. Eine allgemeine Form von Welttechnik würde weiter auf komplexen iterativen Verfahren basieren, die sich imitativ oder selbstverstärkend verhalten. Solche iterativen Verfahren wiederum resultieren in Strukturen, die aufbauend auf Sensordaten bestimmte Eigenschaften entwickeln und dabei zunehmend undurchsichtig beziehungsweise intransparent werden. Diese Art von Design oder Konstruktionsverfahren, die im Rahmen des maschinellen Lernens technisch realisiert ist, kann durchaus auch in anderen Ingenieurszusammenhängen auftreten¹⁵⁴. Die nächste Aufgabe besteht entsprechend darin, konkrete iterative Verfahren der genannten Art zu identifizieren und sie auf ihre Vergleichbarkeit mit maschinellem Lernen hin zu betrachten.

Ein Beispiel für einen Bereich von Technik, in dem die Grundoperationen der Steuer- und Regelungsprozesse ebenfalls nicht immer realisiert werden können, bildet die Nanotechnik. Hier werden ebenfalls mitunter mittels medialer Steuerung Umgebungen inszeniert um zu prüfen, welche Phänomene und Strukturen unter speziellen Rahmenbedingungen entstehen können. Der Hilfsbegriff zur Beschreibung dieser Vorgehensweise ist der der Inszenierung, der nicht zufällig auch bei Rheinberger Verwendung findet. Zwar wäre ein großer Schritt getan, wenn die Nanotechnik sich als Welttechnik verorten ließe, allerdings steht dem ein erhebliches Problem im Weg. *Die Nanotechnik* ist eine gleichermaßen unpräzise Beschreibung wie die Rede von *dem maschinellen Lernen* oder gar *der Informatik*. Hier fehlt eine systematische, interdisziplinäre Darstellung der technischen Grundlagen in Hinblick auf Selbstorganisation, wie sie diese Arbeit für das maschinelle Lernen darstellt. Dasselbe gilt auch für andere Technikbereiche von denen vermutet werden könnte, dass dort eine Suche nach Welttechnik sinnvoll sein könnte. Eine interdisziplinäre Darstellung dieser Technikbereiche kann nicht von den jeweiligen Disziplinen selbst verlangt werden, da aus

¹⁵⁴ Tatsächlich ist gerade das Auftreten solch eines Designs in der Informatik eher kontraintuitiv, da das Auftreten unvorhergesehener Strukturen beim Einsatz von zufallsfreien Algorithmen zunächst überrascht.

einer rein disziplinären Perspektive die interdisziplinären und technikphilosophischen Fragestellungen nicht sehr präsent sind. Soll dementsprechend die Suche nach Welttechnik und die Diskussion und Analyse von Selbstorganisation nicht zurückgestellt werden, bis ein interdisziplinäres Projekt sich mit der relevanten Facette eines weiteren Technikbereiches beschäftigt, so müssen als Suchraum technikphilosophische Perspektiven genutzt werden, die einen größeren Allgemeinheitsanspruch und eine entsprechend verringerte Auflösungskraft aufweisen¹⁵⁵. Den Vorzug können hierbei interdisziplinär geprägte, technikphilosophische Perspektiven erhalten, die zwar einerseits gegebenenfalls nicht ohne Weiteres auf maschinelles Lernen anwendbar sind, aber andererseits auch keine allgemeine Technikphilosophie darstellen, sondern primär Technikbereiche betrachten, die auch im Rahmen der Diskussion von Selbstorganisation betrachtet werden. Ein Beispiel für solch eine Perspektive, das bereits in der vorangegangenen Herleitung von Welttechnik betrachtet wurde, ist Rheinbergers Darstellung epistemischer Dinge. Im Weiteren werden in Kürze zwei weitere Perspektiven benannt, die die genannten Kriterien erfüllen und für eine weiterführende Analyse des maschinellen Lernens herangezogen werden könnten.

4.1.1. Technoscience als möglicher Suchraum

Eine noch nicht genannte Perspektive, die einen besonders vielversprechenden Suchraum für weitere Beispiele von Welttechnik darstellt, bietet die **Technoscience**. Die Technoscience betrachtet Dinge explizit als an der Erzeugung von Möglichkeiten und Weltbezügen beteiligt.

Things demand intimacy, things are partners that afford possibilities, create successful connections to the world.

(Nordmann et al. 2011, S. 2)

Gerade wenn anerkannt wird, dass spezielle Teile des maschinellen Lernens sich deutlich von der klassischen Technik unterscheiden, könnte versucht werden die Überlegungen zum Status von Nanotechnik als einer Technoscience auch auf MLA und auf Welttechnik abzubilden. In der Perspektive der Technoscience auf Dinge – wie eben auch Strukturvorschläge – ist insbesondere die Notwendigkeit des In-ein-Verhältnis-Setzens mit den Dingen mitgedacht.

Affordances – a relational property that becomes salient only in respect to human purposes and material agencies.

(Nordmann et al. 2011, S. 2)

Zwar kreieren mit Hilfe von Welttechnik erstellte Strukturvorschläge keine Verbindungen zur Welt, sondern basieren auf dem nutzerseitigen Wunsch nach einer Ersterstellung von Weltbezügen, aber diese Differenz scheint überbrückbar. Ein grundsätzlicheres Problem und gleichzeitig implizit ein konstruktiver Umgang mit dem Problembegriff der Selbstorganisation findet sich in der folgenden Kurzdefinition von Technoscience.

¹⁵⁵ Eine Steigerung der Komplexität der technischen Seite der Betrachtung, etwa ein Vergleich von Nanotechnik oder Genetik mit maschinellern Lernen, würde noch höhere Ansprüche stellen als eine direkte Analyse der Nanotechnik. Diese Steigerung der Komplexität wird schon daraus deutlich, dass nicht klar ist, was ein dem maschinellen Lernen als Teilbereich der Informatik vergleichbarer Teilbereich der Nanotechnik wäre.

Technoscience – the knowledge-production of homo faber that often uses scientific representations (e.g. theories, models, diagrams) to make things work.

(Nordmann et al. 2011, S. 1)

In Hinblick auf Welttechnik ist fraglich, wie notwendig es ist, dass gerade einem neugierigen Umgang mit Dingen wirklich wissenschaftliche Repräsentationen zugrunde liegen.

Eine Aufarbeitung der Potenziale und Schwierigkeiten der Suche nach Welttechnik mit Hilfe der Perspektive der Technoscience kann und soll im Rahmen dieser Analyse nur angedeutet werden. Gleichwohl kann festgehalten werden, dass die Diskussion der Technoscience stark interdisziplinär geprägt ist und den Dialog mit techniknahen Experten sucht. Diese Ausrichtung und die technikleiche Argumentation lassen die Technoscience als einen besonders vielversprechenden Suchraum für die Identifikation von Welttechnik erscheinen.

4.1.2. Abduktionsstufen als Mittel zur Suche

Eine zweite Möglichkeit, nach dem Auftreten von Welttechnik zu suchen, basiert auf der Beschreibung des Agierens von Artefakten als der Erstellung logischer Schlüsse und besteht in der Analyse und Einordnung der jeweiligen Schlussweisen.

This study embeds abductive search for new ideas within a broader framework concerning human cognitive activity than just reasoning as such.

[...]

Instead of conceptualizing human beings as processing information just in their heads inquirers are seen as fundamentally connected both to other inquirers (cultural resources, communities) and nature (world, indexical object of inquiry) through developing mediating artifacts and shared objects of activity collaboratively.

(Paavola 2006, S. 70)

Die im ersten Hauptteil vorgenommene Darstellung von Konzepten zur Erstellung von Autoadaptionsprozessen reicht aus, um die Möglichkeit einer pauschalen Antwort auf die Frage nach der Schlussweise von Autoadaptionsprozessen als unwahrscheinlich und die Frage somit als durchaus interessant erscheinen zu lassen. Im Rahmen der Darstellung des analytischen Lernens wurde bereits angedeutet, dass die Frage nach der logischen Schlussweise von MLA auch innerhalb der Informatik gestellt wird. Die Begriffsverwendungen sind jedoch sehr uneinheitlich und zum Teil widersprüchlich¹⁵⁶ (Flach et Kakas 2000; Kaminski et Harrach 2010).

Der Klärung dieser Begriffsverwendungen und der anschließenden Nutzung der Perspektive bei der Suche nach Welttechnik steht zunächst eine Gefahr im Wege, die ähnlich gelagert ist

¹⁵⁶ Eine Verbindung der Perspektive von Peirce und der Nutzung des Abduktionsbegriffes (Peirce 1878; Peirce et Walther 1991) bieten Costa und Wirth (Costa 2008; Wirth 2003, S. 19). Einen Versuch der Darstellung der Konsequenzen der unscharfen Begriffsbildung bietet Gottschalk (Gottschalk 1999).

wie die Schwierigkeiten bei einer verstärkten Nutzung des Begriffes des Lernens. In beiden Fällen scheinen sich sofort starke Parallelen zwischen MLA und Menschen aufzudrängen.

Der Versuch, den Denkprozeß in Analogie zum Verarbeitungsprozeß des Computers zu setzen, führt fast zwangsläufig zu der Frage nach der möglichen Funktion der Abduktion innerhalb automatischer Informationsverarbeitung.

(Wirth 1995, S. 14)

Auch bei einer Betrachtung der Rede von Abduktion muss die Analyse sich dementsprechend dagegen absichern, Artefakten implizit menschliches Verhalten zu unterstellen. Die in der Welttechnik angelegte Trennung der Beiträge des unterstützenden Artefaktes und des Nutzers – der Präsentation und der Interpretation – bei der Erstellung von Weltbezügen lässt sich bei der Betrachtung des Autoadaptionsprozesses als eines logischen Schlusses nicht ohne Weiteres aufrechterhalten. Die Erstellung eines Strukturvorschlages, der eine gewisse Fähigkeit entwickelt, stellt nicht unbedingt einen vollständigen abduktiven Schluss gleich welcher Stufe dar. Die Beschreibung, dass die Eigenschaften von Strukturvorschlägen genutzt werden können, um menschliche abduktive Schlüsse zu ermöglichen und zu unterstützen¹⁵⁷, kann als Reformulierung eines zentralen Aspektes von Welttechnik verstanden werden, muss jedoch nicht dazu führen, den noch nicht interpretierten Strukturvorschlag als das Ergebnis eines logischen Schlusses anzusehen. Dies gilt vor allem, wenn die zugrunde liegende Lernstrategie sehr stark neugierig geprägt ist und eine stark autoadaptive Vorstruktur nutzt, wodurch der Strukturvorschlag gerade keine klare Menge von Entscheidungsregeln darstellt.

Dennoch lassen sich die Vorgehensweisen von MLA bei einer impliziten Mitbetrachtung des menschlichen Anteils durchaus sehr produktiv als verschiedene Stufen abduktiver Schlüsse beschreiben (Kaminski et Harrach 2010). Insbesondere erleichtert es diese Perspektive, den eigentlichen Autoadaptionsprozess nicht zu Gunsten der Betrachtung des Strukturvorschlages zu vernachlässigen. Die Rede von einem abduktiven Schluss bezieht sich immer auf den Prozess der Erstellung des Strukturvorschlages und berücksichtigt insbesondere die Vorstruktur als Modell für die technischen Formalisierungen der jeweils eingesetzten Selbstorganisationsprinzipien. Gerade die Unterscheidung zwischen neugierigen und zielorientierten Artefakten profitiert von dem Fokus auf die Vorstruktur und der Diskussion, welcher **Stufe** ein abduktiver Schluss zugeordnet werden kann. Abduktive Schlüsse höherer Stufe zeichnen sich dadurch aus, dass im Vorfeld ein geringeres Maß an Informationen bekannt ist.

All other forms [of abductive reasoning] are derived thereafter by removing the elements from the set of information given in the basic form one at a time.

(Kaminski et Harrach 2010, S. 2)

Das bedeutet, zielorientierte Autoadaptionsprozesse würden am ehesten einem abduktiven Schluss entsprechen, der unter einer Reihe von denkbaren Regeln die im konkreten Kontext

¹⁵⁷ Menschliche Schlüsse können jedoch weiterhin nur in sehr optimierungsnahen beziehungsweise zielorientierten Ausnahmefällen vollständig ersetzt werden.

optimale Regel identifiziert und zum Einsatz bringt. Neugierige Autoadaptionsprozesse hingegen kommen in erster Linie in sehr vage bekannten Kontexten zum Einsatz, in denen eine mathematische Modellierung und Optimierung noch nicht möglich ist und eine autoadaptive Vorstruktur mit geringen Vorgaben zu Vorwissen erstellt werden muss.

4.1.3. Mehrwert einer weiteren Verallgemeinerung

Die hier angedeutete Diskussion von Methoden, die bei der weiteren Suche nach Welttechnik hilfreich sein könnten, zielte darauf ab, eine Begriffsbildung oder Perspektive zu finden, die einerseits ein größeres Feld von Technik betrachtet als nur die Informatik und die andererseits zumindest einige der wesentlichen Teilaspekte der Idee von Welttechnik oder neugieriger Artefakte thematisiert. Eine Vermutung bestand darin, dass eine solche Begriffsbildung aufgrund ihres allgemeineren Anspruches kein ausreichend großes Auflösungsvermögen aufweist, um bestimmen zu können, welche Teile des maschinellen Lernens als Welttechnik identifiziert werden können, dass jedoch der Gewinn von neuen oder präziser gefassten Suchräumen diese Schwäche aufwiegt. Die der Formulierung dieser Schwäche entgegengesetzte Denkrichtung wurde bisher nur implizit mitbetrachtet. Gemeint ist die Frage, ob die Möglichkeit, eine Diskussion maschinellen Lernens führen zu können, nicht generell einen optionalen Charakter hat, wenn andere Technikbereiche betrachtet werden sollen und bereits festgestellt wurde, wie maschinelles Lernen zu beschreiben ist. Weitergedacht kann sogar gefragt werden, ob schon die in dieser Analyse gewonnene Perspektive auf Welttechnik überhaupt noch in der Lage ist, zwischen den im ersten Hauptteil vorgestellten Lernstrategien zu unterscheiden. Tatsächlich ist die Perspektive der Welttechnik hierzu *nicht* in der Lage, denn die Betrachtung der Welttechnik setzt andere Schwerpunkte. Die Unterscheidung zwischen zielorientierten und neugierigen Artefakten etwa hängt nur lose mit der technischen Unterscheidung zwischen einzelnen Lernstrategien zusammen.

Der Analyseschwerpunkt im zweiten Hauptteil wurde bewusst von der Klassifizierung des ersten Hauptteils gelöst und diese Flexibilität stellt auch kein Problem dar, denn der erste Teil sollte gerade einen interdisziplinären und technikphilosophischen Diskurs darüber ermöglichen, was an maschinellem Lernen interessant sein könnte. Entsprechend kann festgehalten werden, dass die Freiheiten bei einer Übertragung der Rede von Welttechnik auf andere Kontexte größer ist, als es im Kontext dieser techniknahen Grundlagenarbeit zu maschinellem Lernen zunächst wirkt. Die Fähigkeit zur Auflösung von Algorithmenklassen stellt in anderen Kontexten kein notwendiges Erfolgskriterium dar. Die Darstellung einer interdisziplinären Grundlagenforschung etwa zur Nanotechnik kann und sollte die technischen Konzepte und Intuitionen, die in dieser Analyse der erste Hauptteil vermittelt, komplett neu bestimmen.

Unabhängig davon wäre es speziell für eine Diskussion des maschinellen Lernens außerordentlich hilfreich, bestimmen zu können, was genau die nutzbaren Lernstrategien voneinander unterscheidet und vor allem, welche weiteren, noch nicht entdeckten Möglichkeiten es geben kann, auf Selbstorganisationsprozessen basierende Vorstrukturen zu entwerfen. Die Konzepte der Informatik zum Einsatz von Selbstorganisation spiegeln sich in den Kriterien zur Bewertung der Autoadaptionsprozesse oder Strukturvorschläge wieder. Die Messung von Performanz, Beschreibungslänge, Fitness, Signalstärke, Abständen, Wahrscheinlichkeiten, Aussagenlogik oder formalen mathematischen Fehlertermen setzt jeweils einen speziellen Zugang zu Selbstorganisationsprozessen voraus. Weiter besteht kaum Veranlassung zu vermuten, dass diese Liste und die korrespondierenden Lernstrategien einen

Anspruch auf Vollständigkeit besitzen. Denkanstöße zur Erweiterung dieser Liste könnten durchaus aus einem interdisziplinären Diskurs mit Beteiligung der Technikphilosophie gewonnen werden. Hierzu wäre es erforderlich, andere Kontexte von Welttechnik zu identifizieren und zu beschreiben oder die Arbeit an den Problem Begriffen der Selbstorganisation oder der Neugier weiter voran zu bringen und die Ergebnisse anschließend in den Kontext des maschinellen Lernen zurück zu übersetzten beziehungsweise zu übertragen.

4.2. Technikferne Einordnung von MLA

In den vorangegangenen Betrachtungen wurde Technik stark isoliert von ihrem gesellschaftlichen oder sozialen Kontext betrachtet. Beispielsweise wurde bei der Positionierung zum Begriff des Zweckes danach gefragt, was mit dem Einsatz des MLA erreicht werden soll, ohne zu berücksichtigen, dass Zwecke mehrstufig angeordnet sein können¹⁵⁸ oder dass eine Vielzahl praktischer Erwägungen und externer Einflüsse den Einsatz eines MLA motivieren können. Die Nichtberücksichtigung von Einflüssen dieser Art im Rahmen einer Grundlegung zum maschinellen Lernen soll an dieser Stelle nicht in Frage gestellt werden. Stattdessen soll eine technisch angemessene, interdisziplinäre Diskussion technikferner Themen durch diese Arbeit gerade ermöglicht werden.

Beispielsweise besteht ein Aspekt der Debatte über den sozialen Charakter von Technik, der in verschiedenen Beispielen dieser Arbeit eine Rolle spielte, in der Frage, inwiefern die Nutzer von Last.fm oder die Kunden eines Kaufhauses durch maschinelles Lernen geformt werden. Diese Debatte greift eine besonders interessante Fragestellung auf, da argumentiert werden kann, dass nicht nur MLA in Hinblick darauf entwickelt werden, vorweg zu nehmen was Menschen als potenzielle Nutzer benötigen. Unabhängig davon liegt es in Hinblick auf MLA nahe, explizit die „Formung“ der Nutzer beim Einsatz von zielorientierten, vorstrukturierten MLA und beim Einsatz von neugierigen, präsentierenden Artefakten zu vergleichen. Auch und gerade die Konzeption neugieriger MLA kann es mangels Zielen oder Zwecken mit sich bringen, dass die Nutzerwünsche erst aus einer Interpretation des Strukturvorschlages entstehen und damit aus Sicht des Nutzers in gewisser Weise vorgegeben werden – unabhängig davon, wie das Artefakt die konzeptionierte Aufgabe konkret löst. Auch wenn zielorientierte MLA eingesetzt werden, spielt es eine zentrale Rolle, zu entscheiden welche Konzepte von den Nutzern vorgegeben werden und Teil der Vorstruktur sein sollen und welche Ziele durch die Berücksichtigung der Rohdaten erreicht werden sollen. Die Übergabe von Rohdaten an ein MLA ist dementsprechend noch in viel allgemeinerer Hinsicht problematisierbar als bisher beschrieben. Im Fall der Betrachtung von Last.fm ist es etwa von Interesse, welche Rohdaten von dem Nutzer vorgegeben wurden, der den Strukturvorschlag betrachtet und welche Eingaben von anderen Nutzern oder den Programmierern des MLA stammen.

Wie bereits betont wurde, sollen die Darstellungen der vorliegenden Analyse interdisziplinäre Diskurse des maschinellen Lernens ermöglichen und motivieren. Die entsprechenden Fragestellungen sollen auf Basis dieser Grundlagenforschung unter Vermeidung von technikfernen Pauschalisierungen und komplexitätsreduzierenden Verkürzungen diskutiert werden können.

¹⁵⁸ Dies wurde jedoch mitbetrachtet, als der MLA als ein gestiftetes Worumwillen beschrieben wurde.

4.3. Maschinelles Lernen ohne Computer

Abschließend wird die bisherige Perspektive auf die im maschinellen Lernen zum Einsatz kommenden Formen von Selbstorganisation noch einmal erweitert. Es wird die These vertreten, dass die Diskussion der Konzepte hinter dem maschinellen Lernens nicht erst mit Hilfe der Informatik möglich wurde. Die Implementierung maschinell lernender Algorithmen durch die Informatik hat lediglich die Geschwindigkeit erhöht, mit der Autoadaptionsprozesse umgesetzt werden können.

Diese These basiert auf der Beobachtung, dass gewisse Formen maschinellen Lernens auch als menschliche Praxis ohne Nutzung eines Computers oder eines anderen technischen Systems möglich sind. Ein Beispiel für evolutionäres Lernen *ohne* Nutzung von Computern bildet die Weiterentwicklung einer Düse zur Herstellung von Seifenpulver in einer Fabrik bei Liverpool.

Da [eine mathematische Optimierung] nicht klappte, versuchten [die Ingenieure] es auf andere Weise [...] günstige Abweichungen wurden beibehalten, ungünstige verworfen. Mit anderen Worten, es handelt sich um natürliche Selektion.

[...]

Man nehme eine Düse, die schon ganz gut funktioniert, und stelle Kopien davon her, die jeweils nach dem Zufallsprinzip ein wenig abgewandelt sind. Dann prüfe man, wie gut sie das Pulver herstellen. Anschließend zwingt man ihnen einen Kampf ums Dasein auf, indem man dafür sorgt, daß nicht alle überleben können. Viele der abgewandelten Apparate sind nicht besser (sondern oft sogar schlechter) als die Ausgangsform. Sie werden verworfen, und nur die wenigen, die bei der Erfüllung ihrer Aufgabe überlegen sind, dürfen sich fortpflanzen. Im Laufe der Generationen schält sich wie von Zauberhand ein neues, hervorragend funktionierendes Rohr heraus, das eine komplizierte, überraschende Form besitzt.

(Jones 2002, S. 119)

Sofern der Selektionsprozess nicht gezielt beeinflusst wurde, entsprechen die von den Ingenieuren auf diese Weise erstellten Düsen neugierig erstellten Strukturvorschlägen. Die Aussagen, dass eine mathematische Optimierung im Vorfeld nicht möglich war¹⁵⁹ und dass die Ergebnisse eine Überraschung darstellten, deuten darauf hin, dass solche Beeinflussungen tatsächlich nicht vorgenommen wurden. Die Rede von einer überraschenden Form legt zwar nahe, dass bezüglich des Aussehens der Düse Erwartungen bestanden, aber es kann vermutet werden, dass diese Erwartungen den Adaptionsprozess nicht beeinflusst haben.

Ein Beispiel für den Einsatz von Data Mining ohne Computer bildet das **Music Genome Project** und, mit Einschränkungen, dessen kommerzielle Nutzung im Rahmen des Last.fm-Konkurrenten **Pandora**. Das Music Genome Project stellt eine Datenbank dar, die Musiktiteln mit 150 bis 400 bewerteten Attributen eine vergleichsweise große Zahl von

¹⁵⁹ Die Düse diente dem Zerstäuben einer Flüssigkeit und die Berechnung des Verhaltens von Flüssigkeiten ist mathematisch extrem anspruchsvoll, wenngleich bei der Modellierung und Simulation inzwischen Fortschritte gemacht werden (Klöppel et al. 2011).

Eigenschaften zuordnet. Das Besondere an dieser Datenbank ist, dass die Attribute vollständig von menschlichen Experten zugeordnet und bewertet werden.

Every time a new song comes out, someone on Pandora's staff — a specially trained musician or musicologist — goes through a list of possible attributes and assigns the song a numerical rating for each one. Analyzing a song takes about 20 minutes.

The people at Pandora — no relation to the alien planet — analyze 10,000 songs a month. They've been doing it for 10 years now, and so far they've amassed a database containing detailed profiles of 740,000 different songs. Westergren calls this database the Music Genome Project.

(Grossmann 2010)

Die von den Mitarbeitern von Pandora erstellten Strukturvorschläge, in Kombination mit einer analytisch im Vorfeld bestimmten mathematischen Modellierung des Zusammenhangs der Attribute, stellen eine typische Anwendung maschinellen Lernens im Data Mining dar, verwenden jedoch erst im kommerziellen Online-Angebot von Pandora MLA. Pandora stellt den registrierten Kunden zusätzlich zu den Experteneinschätzungen und analog zu Last.fm auch Strukturvorschläge basierend auf Empfehlungen von anderen Nutzern zur Verfügung. Welche Attributausprägung ein Nutzer konkret bevorzugt, muss im Einzelfall festgestellt werden, aber diese Analyse entspricht einerseits einem nachgelagerten Schritt und kann andererseits relativ einfach über die Beantwortung eines umfangreichen Fragebogens festgestellt werden. Die kommerzielle Natur von Pandora verlangt hier aber naheliegenderweise nach einer komfortableren Lösung, die in Form von mathematischer Modellierung oder eben maschinellern Lernen zur Verfügung gestellt wird.

Die Vorgehensweise des Music Genome Projects stellt einen extrem zielorientierten Prozess dar, der sehr langfristig geplant werden musste, um die Verfügbarkeit von ausreichend vielen Experten sicherzustellen. Das Music Genome Project lässt sich sehr gut von der Webseite Pandora trennen¹⁶⁰ und mit dem isoliert betrachteten Music Genome Project liegt ein beeindruckendes Beispiel für Data Mining ohne Einsatz eines maschinell lernenden Artefaktes vor. Weiter stellt das Projekt durch seine Kopplung an Pandora eine besonders transparente Veranschaulichung dafür dar, dass häufig verschiedene Arten maschinellen Lernens im Zusammenspiel mit anderer Technik eingesetzt werden.

Der Mehrwert, den die Informatik in solchen Methodenkompositionen bietet, liegt darin, dass Selbstorganisationsprozesse und Datenbankerstellung automatisch, extrem schnell und mit klar benennbaren systematischen Fehlern durchgeführt werden¹⁶¹. Der Einsatz von leistungsstarken Computern erlaubt das Entstehen von Welttechnik, deren neugierige Strukturvorschläge unüberschaubare Datenmeere erst wahrnehmbar werden lassen.

¹⁶⁰ Wohl auch, damit gezielt Marketing mit dem scheinbar unkommerziellen Music Genome Project betrieben werden kann.

¹⁶¹ Im Gegensatz zu den weniger transparenten systematischen Fehlern, die ein menschlicher Experte macht.



Abbildungsverzeichnis

Abbildung 1:	Beispiel für Pseudocode.....	16
Abbildung 2:	Überblick der Möglichkeiten zur Unterscheidung von MLA.....	18
Abbildung 3:	Beispielverteilung von Würfelergebnissen eines asymmetrischen Würfels.....	20
Abbildung 4:	Überblick der Möglichkeiten stochastischer Suche.....	24
Abbildung 5:	Auflösung des maschinellen Lernens nach Lernstrategien.....	27
Abbildung 6:	Entscheidungsbaum zur Wahl des Mittagessens	29
Abbildung 7:	Trainingsdaten des Beispiels zur Kreditwürdigkeit.....	35
Abbildung 8:	Ungestutzter Entscheidungsbaum.....	35
Abbildung 9:	Gestutzter Entscheidungsbaum	36
Abbildung 10:	Strukturvorschlag in Form eines Regressionsbaumes	37
Abbildung 11:	Evolutions- und Selektionsschritt am Beispiel geometrischer Objekte.....	41
Abbildung 12:	Grundlage einer Beispielcodierung von „Kanu fahren“	44
Abbildung 13:	Auswertungsbaum beziehungsweise Parse Tree der Formel „A + 2B“	45
Abbildung 14:	Aufbau und Zielzustand des Turmbaus von Hanoi	47
Abbildung 15:	Notationsbeispiel eines Vektors mit zufälligen Werten	49
Abbildung 16:	Unterschiedliche Autoadaptionszyklen im evolutionären Lernen.....	50
Abbildung 17:	Vollständig verbundenes künstliches neuronales Netz.....	57
Abbildung 18:	Zwei Möglichkeiten einer einfachen Signalgewichtung	58
Abbildung 19:	Backpropagation als Aktualisierungsregel überwachten Lernens	60
Abbildung 20:	Selbstorganisierende Karte, die eine Eingabe der Kategorie I zuordnet	61
Abbildung 21:	Verbindungsgewichte als Werte der Raumachsen Höhe, Breite und Tiefe..	61
Abbildung 22:	Adaption von Verbindungsgewichten mittels eines Eingangssignals	62
Abbildung 23:	Übersicht der Schichten eines dreischichtigen KNN.....	64
Abbildung 24:	Lineare Trennung (links) und nicht-lineare Trennung (rechts)	65
Abbildung 25:	Realisierung der logischen Funktion UND mittels Verbindungsgewichten .	65
Abbildung 26:	Rekursive Verbindung in einem rekurrenten KNN.....	67
Abbildung 27:	Voronoi Diagramm mit fünf Teilflächen.....	71
Abbildung 28:	Voronoi Diagramm mit einer Nachbarschaft	73

Abbildung 29:	Vereinfachung eines Voronoi Diagramms.....	77
Abbildung 30:	Histogramm in Form einer Normalverteilung.....	79
Abbildung 31:	Bedingte Wahrscheinlichkeit des Antreffens von Vegetariern	81
Abbildung 32:	Bedingte Wahrscheinlichkeiten mit mehreren Attributen	83
Abbildung 33:	Trainingsdaten in Form von geometrischen Objekten	84
Abbildung 34:	Bayes'sches Netz mit Wahrscheinlichkeitstabellen.....	86
Abbildung 35:	Bayes'sches Netz mit zwei Einflussgrößen.....	87
Abbildung 36:	Stammbaum als Einführungsbeispiel zur ILP.....	92
Abbildung 37:	Fiktive Grafik zum Entgelt von Angestellten.....	95
Abbildung 38:	Lineare Trennung mit kleiner (I) und großer (II) neutraler Zone	96
Abbildung 39:	Übersicht der potenziellen Problembegriffe.....	125
Abbildung 40:	Beispiele für naturalisierte Technik.....	142
Abbildung 41:	Raum naturalisierter Technik – Visualisierung A.....	143
Abbildung 42:	Raum naturalisierter Technik – Visualisierung B	145
Abbildung 43:	Raum naturalisierter Technik – Visualisierung C.....	147
Abbildung 44:	Raum naturalisierter Technik – Visualisierung D.....	147
Abbildung 45:	Grenzübergänge im Raum naturalisierte Technik.....	148
Abbildung 46:	Raum transklassischer Technik	153
Abbildung 47:	Einfache und komplexe Autoadaption im evolutionären Lernen.....	163
Abbildung 48:	Geschätzte Häufigkeiten des Auftretens von MLA	165

Stichwortverzeichnis

Abduktionsstufe	200	Bayes-Klassifikator, optimaler	89, 91
Abduktive Verzerrung	<i>Siehe</i> Bias	Bayes'sches Lernen.....	80
Abhandenheit.....	133	Bayes'sches Netz	80, 86
Adaptive Resonanztheorie	62	Bewandtnisganzheit.....	121
Ähnlichkeit.....	21	Bias.....	33
Aktivierung	57	Biofakt.....	158
Aktualisierungsregel.....	59	Bitfolge.....	43
Algorithmus	115	Black Box.....	100
Fauler Algorithmus	71	Blatt.....	30
Optimierungsalgorithmus	46	Brute-Force	21
Als-frei	131	Changieren	119
Ambient Intelligence	17	Cluster.....	21, 74, 186
Analytisches Lernen.....	28	Clusteranalyse	21, 74, 88
Anfangspopulation	39	Agglomerative Clusteranalyse.....	74
Angemessenheit	180, 186	Hierarchische Clusteranalyse.....	77
Artefakt.....	3	Codierung.....	40
Assoziationslernen.....	80, 83	Crowding.....	50
Assoziationsregel.....	21, 82, 83	Cyberfakt.....	158
Ast	33	Darstellungskraft	63
Atomar.....	43	Daseinsanalytik.....	118
Attribut	28	Data Mining.....	2, 179
Ausreißer	76	Daten	
Autoadaptionsprozess.....	3, 8	Eingabedaten	28
Instanzenbasierter Autoadaptionspr....	75	Rohdaten	6, 18
Konzeptorientierter Autoadaptionspr. .	75	Testdaten	36, 76
Autoadaptionsstrategie.....	18	Trainingsdaten	18, 30
Backpropagation	59, 69	Validierungsdaten	36, 76
Baldwin Effekt.....	51	Decodierung	40
Baum	45	Deduktives Lernen	91, 93
		Deskriptives Lernen	69
		Eigentätigkeit.....	122

Einflussfaktor	86	Logische Funktion	65
Einfrieren	29	Transformationsfunktion	161
Einheit	56	Zielfunktion	48
Ausgabeeinheit	57	Gegebenheit.....	148
Eingabeeinheit.....	57	Generation.....	38
Sigmoide Einheit	65	Genetische Programmierung	45
Verborgene Einheit	57	Genetischer Algorithmus.....	43
Elter	49	Genotyp.....	45
EM-Algorithmus	88	Genotyp-Phänotyp-Abbildung	43
Entropie	31, 107	Gewicht	56, 58
Entscheidungsbaum	27, 28, 107	Verbindungsgewicht	58
Entscheidungswald.....	36	Gewinnverhältnis.....	32
Enttäuschung	106	Gliazelle.....	53
Epistemische Dinge	111	Go	194
Epoche	58	Graph	
Erklärungs-basierte Lernen.....	94	Azyklischer Graph	86
Erwartbarkeit	105, 186	Grundlagenforschung	14
Erwartung	107, 114, 116, 128, 138, 162	Halbzug	100
Höherstufige Erwartung.....	109	Handeln	
Erwartungswert.....	80	Instrumentelles Handeln	100
Evolutionärer Operator.....	40, 163	Hebb'sche Lernregel.....	52
Evolutionäres Lernen.....	27, 38, 107, 138	Heuristik.....	93
Expectation-Maximization-Algorithmus ..	88	Hinwegerklären	88
Experiment.....	111, 179	Hyperebene	95
Fallbasiertes Lernen	72	Hypothesenlernen	
Fehler.....	59	Deskriptives Hypothesenlernen.....	73
Fitnessfunktion.....	40	Globales Hypothesenlernen	73
Flow	17	Lokales Hypothesenlernen.....	73
Fragmentierung.....	32	Prädikatives Hypothesenlernen.....	73
Funktion	8, 30, 65	ILP	91
Aktivierungsfunktion	57	Implementierung	3
		In-der-Welt-sein.....	121

Individuum	38	Laufzeit.....	33, 34
Induktion	123, 124, 175	Lernen	
Induktive logische Programmierung	91	Bestärkendes Lernen	21
Information.....	31	Überwachtes Lernen.....	19
Informationsgewinn.....	32	Unüberwachtes Lernen.....	20
Informationstechnik.....	115, 158, 185	Lernstrategie.....	11, 15, 18, 25
Informationstrennung.....	32	Nachgeordnete Lernstrategie.....	75
Instanz	30	Linearität	64
Testinstanz	18	Maschinell lernendes Artefakt.....	3, 7
Trainingsinstanz	18	Maschinelles Lernen.....	1
Instanzenbasiertes Lernen	28, 138	Maschinenkomponente	161
Inszenierung	108	Maske	44
Intervall	36	Maximierung	48
Kante	30	Medialität	152
Klassifikator	29, 97	Medienphilosophie.....	123
Klassifizierung.....	21, 29	Mischungszahl	49
Dynamische Klassifizierung.....	30	MLA.... <i>Siehe</i> Maschinell lernendes Artefakt	
Statische Klassifizierung.....	30	Modell	
KNN	27, 51, 99, 115, 138	Globales Modell	37
Azyklische KNN	63	Lokales Modell	21
Rekurrente KNN	63, 66	Moment.....	121
Knoten	30	Multi-Rekombination	49
Kinderknoten.....	31	Music Genome Project	204
Wurzelknoten	30	Mutation.....	40
Kompetitives Lernen.....	<i>Siehe</i> Wettbewerbslernen	Mutationsschrittweite.....	48, 150
Konnektionismus.....	52	Mutationsschrittweitenband	48
Konvergenz	48	Nachbarschaft.....	72
Konvergierende Technologien	13, 140	Nanotechnik	108
Künstliches neuronales Netz.....	<i>Siehe</i> KNN	Naturalisierte Technik.....	141
Kunst	116, 171, 175	NBIC.....	140, 160, 188
Kunstwerk	150	Nervenzelle.....	52

Neugier	125	Pervasive Computing	142
Neugierige Artefakte	130, 133, 138	Perzeptron	
Neuron.....	52, 57	Einlagiges Perzeptron	64
Künstliches Neuron.....	57	Zweilagiges Perzeptron	65
Nichterwartbarkeit	106, 116	Phänotyp	45
Nichttriviale Maschine.....	160	Population	38
Nichttrivialität		Präsentation.....	107
Nichttriviale Maschine	161	Probabilistisch.....	41
Nichttriviale Nichttrivialität erster		Problemstellung.....	30
Ordnung	164	Pseudocode.....	15
Nichttriviale Nichttrivialität zweiter		Qualität	30, 31
Ordnung	164	Quantenphysik.....	136
Nichtwissen.....	13, 152, 186	Raum	
Nullzustand.....	52	Gewichtungsraum	58
Nutzer	7	Hypothesenraum	22
Nutzeroberfläche	7	Lösungsraum.....	4, 22
Nutzerstereotyp	156	Suchraum.....	22
Objektparameter	49	Rauschen	33
OLAP-Würfel	2	Recommender-System	2, 7, 26
Optimierung.....	48, 95, 97, 128	Regel	91
Optimierungsaufgabe.....	38	Regelung	104, 108, 152, 185
Organismus	13	Regressionsbaum	36
Pandora	204	Rekombination	40
Parallelkommunikation	156	Relevanzbasiertes Lernen.....	93
Parallelverarbeitung	52	Repräsentant	30, 37
Parameter	20, 80	Repräsentation.....	40, 107
Verborgener Parameter.....	75	Lokale Repräsentation	63
Wachsamkeitsparameter	62	Repräsentationsoptionen	109
Parametrisierung.....	50, 125	Resonanz	62
Parse Tree	45	Richtigkeit	180
Partitionierung	13, 32	Robustheit	33
Performanz	34	Satz von Bayes.....	82

Schätzer	80	Nichtparametrischer Ansatz.....	80
Schätzung	78	Parametrischer Ansatz.....	80
Schicht	64	Semiparametrischer Ansatz	80
Ausgabeschicht	64	Stellung	54
Verborgene Schicht.....	64	Stetigkeit	36
Schnittstelle	155	Steuerung	104, 108, 152, 185
Schwellenwert.....	57	Mediale Steuerung	108, 111
Selbstorganisation	1	Stichprobe	23, 79
Selbstorganisierende Karte	60, 132	Repräsentative Stichprobe	78
Selektion		Stiftung.....	127
Roulette Selektion	46	Stochastik	30
Turnierselektion	46	Strategieparameter	49
Sensordaten	1	Endogener Strategieparameter	49
Separierung.....	115	Exogener Strategieparameter	49
Sequenz	21	Strukturvorschlag	6, 22, 26, 32
SETI.....	187	Neugierige Strukturvorschläge	1, 205
Sich-in-ein-Verhältnis-Setzen	123, 160, 185	Stützvektor	96
Signal.....	56	Stützvektormethode	28, 70, 94, 138
Signalfluss	59	Stutzung	33
Simulation	111, 157	Regelstützung	34
Smart	17, 161	Subgruppe	74
Smarte Artefakte.....	118	Subgruppenentdeckung.....	74
Speicherplatz	34	Suche.....	97
Spezifikation	39	Geordnete Suche	23
Spur	135	Gradientensuche	23
Stabilität	109	Semiparametrische Suche.....	24
Starke Kausalität	48	Stochastische Suche	23
Statistik.....	79	Strahlensuche.....	42
Statistische Inferenz	79	Unvollständige Suche	33
Statistisches Lernen.....	28, 70, 138	Ziellose Suche	104
		System	
		Dynamisches System	66
		Subsymbolisches System	29
		Symbolisches System.....	29, 170

Technikkompetenz	156	Vor-Struktur	125, 134, 156, 162
Test	30	Wahrscheinlichkeit	
Theoretische Informatik	15	Bedingte Wahrscheinlichkeit	81
Toy Modeling	108	Unbedingte Wahrscheinlichkeit	81
Training	108	Wahrscheinlichkeitstabelle	86
Transklassische Technik	13, 152, 185	Warenkorbanalysen	82
Trennung	31	<i>Ways of Worldmaking</i>	169
UbiComp	<i>Siehe Ubiquitous Computing</i>	Welt	108, 119
Ubiquitous Computing	4, 17, 118	Welttechnik	1, 11, 98, 158, 187
Überanpassung	33	Weltversion	169
Überraschung	106	Wettbewerbslernen	76
Unabhängigkeit	89	Wissenslevel-Lernen	93, 94
Bedingte Unabhängigkeit	89	Worumwillen	
Unfall	113	Gestiftetes Worumwillen	127
Un-Vorher-Gesehen	131	Zeug	121
Varianz	80	Zielfunktion	19
Vektor	49	Ziellosigkeit	104
Verbindung		Zielorientiertheit	46, 121
Gerichtete Verbindung	56	Zielorientierte Artefakte	130
Verteilte Berechnung	52	Zielorientiertes Lernen	137
Verteilung	19, 23, 74, 78	Zufallsexperiment	78, 79
Gleichverteilung	19	Zufallstechnik	112
Normalverteilung	79	Zufallsvariable	86
Vorhabe, Vorsicht und Vorgriff	135	Zuhandenheit	121
Voronoi-Diagramm	71	Zweck	102
		Innerer Zweck	100

Literaturverzeichnis

- Wikipedia (2001). Florida: Wikimedia Foundation, Inc.
- Wikibooks (2003). Florida: Wikimedia Foundation, Inc.
- Aguilar, Carlos; Lipson, Hod (2008): "A robotic system for interpreting images into painted artwork". Cornell Computational Synthesis Lab, Cornell University.
- Alpaydin, Ethem (2008): *Maschinelles Lernen*. München: Oldenbourg.
- Althoff, Klaus-Dieter; Weiß, Stefan (1991): „Fallbasiertes Problemlösen in Expertensystemen. Begriffliche und inhaltliche Betrachtungen“. Kaiserslautern: Fachbereich Informatik, Univ. Kaiserslautern.
- Bauer; Benenson; Westfall (1999): „cliXX Physik - Anzahl der Atome“. Hg. v. Verlag Harri Deutsch.
- Bense, Max; Walther, Elisabeth (1998): *Ausgewählte Schriften. Band 4: Poetische Texte*. Stuttgart [u.a.]: Metzler.
- Bergman, Mats; Paavola, Sami (2003-2011): "The Commens Dictionary of Peirce's Terms. Peirce's Terminology in His Own Words".
- Branavan, S.R.K; Silver, David; Barzilay, Regina (2011): "Learning to Win by Reading Manuals in a Monte-Carlo Framework". Computer Science and Artificial Intelligence Laboratory - Massachusetts Institute of Technology / Department of Computer Science - University College London.
- Brause, Rüdiger (2010): „Adaptive Systeme“. Vorlesung. Universität Frankfurt.
- Brown, Henry; Haggerty, Terry; Hsiao, Gilbert; Melini, Douglas; Zahn, Michael (2007): "Group Exhibition Machine Learning". New York Foundation for the Arts.
- Burkhard, Hans-Dieter (2010): „Moderne Methoden der KI Maschinelles Lernen“. Humboldt Universität zu Berlin.
- Chang-Claude, Jenny; Hermann, Silke; Eilber, Ursula; Steindorf, Karen (2005): "Lifestyle Determinants and Mortality in German Vegetarians and Health-Conscious Persons. Results of a 21-Year Follow-up". In: *Cancer Epidemiol, Biomarkers & Prevention* (14(4)), S. 963–968.
- Costa, Darin McNabb (2008): "Can Creativity Be Formalized? Peircean Reflections on the Role of Abduction in Human Intelligence". In: Rene V. Mayorga und Leonid I. Perlovsky (Hg.): *Toward Artificial Sapience*. London: Springer, S. 3–14.
- Doersch, Carl; Singh, Saurabh; Gupta, Abhinav; Sivic, Josef; Efros, Alexei (2012): "What Makes Paris Look like Paris?" In: *ACM Transactions on Graphics (SIGGRAPH 2012)* (volume 31, No. 3).
- Doggers, Peter (19.04.2009): "Chess Grandmaster Boris Spassky on Computers in Chess". Interview mit Boris Spassky. Nalchik.
- Doggers, Peter (11.11.2010): "The big Dvoretzky interview". Interview mit Mark Dvoretzky. Moskau.
- Doggers, Peter (06.06.2012): "I was by no means inferior in this match". Interview mit Boris Gelfand. Rishon-le-Zion (via Skype).
- Drepper, Johannes (2010): „Die Beteiligung des menschlichen Kleinhirns an kognitiven assoziativen Lernvorgängen“. Bochum.

-
- Dugas, C.; Bengio, Y.; Chapados, N.; Vincent, Pascal; Denoncourt, G.; Fournier, C. (2003): "Statistical Learning Algorithms Applied to Automobile Insurance Ratemaking". In: L. Jain und A.F. Shapiro (Hg.): *Intelligent and Other Computational Techniques in Insurance. Theory and Applications*. World Scientific.
- Dwyer, Douglas (2005): "Examples of overfitting encountered when building private firm default prediction models".
- Eco, Umberto (1983): "Horns, Hooves, insteps: Some hypotheses on three types of abduction". In: Umberto Eco und Thomas A. Sebeok (Hg.): *The Sign of three. Dupin, Holmes, Peirce*. Bloomington: Indiana University Press (Advances in semiotics).
- Einstein, Albert; Heisenberg, Werner (1926): "Einstein objecting to the placing of observables at the heart of the new quantum mechanics". Heisenberg Vorlesung. Berlin, 1926.
- Fahlman, Scott E.; Lebiere, Christian (1990): "The cascade-correlation learning architecture". In: *Advances in neural information processing systems* 2, S. 524–532.
- Fischer, Bobby (1960): "Lessons from Bobby Fischer. Fischer, R. - Euwe, M. (Leipzig 01) [B13] endgame". Hg. v. Exeter Chess Club.
- Fisk, Donald (2006): "Beer and Nappies - A Data Mining Urban Legend". Zu finden auf web.onetel.net.uk.
- Flach, Peter (1994): *Simply logical. Intelligent reasoning by example*. Chichester, New York: Wiley.
- Fleisch, Elgar; Mattern, Friedemann (2005): *Das Internet der Dinge. Ubiquitous Computing und RFID in der Praxis : Visionen, Technologien, Anwendungen, Handlungsanleitungen*. Berlin: Springer.
- Foerster, Heinz von (1993): „Prinzipien der Selbstorganisation im sozialen und betriebswirtschaftlichen Bereich“. In: Heinz von Foerster (Hg.): *Wissen und Gewissen. Versuch einer Brücke*. Frankfurt a. M.: Suhrkamp, S. 233–268.
- Foerster, Heinz von (2003): "For Nicholas Luhmann: How recursive is communication (1993)". In: Heinz von Foerster (Hg.): *Understanding understanding. Essays on cybernetics and cognition*. New York: Springer.
- Fraser, Neil (2003): "Neural Network Follies". Zu finden auf neil.fraser.name.
- Friedel, Frederic (13.03.2006): "Kramnik on health, plans – and computers". Interview mit Vladimir Kramnik.
- Friedman, Nir (1998): "The Bayesian Structural EM Algorithm. Computer Science Division, University of California".
- Gloy, Karen; Zur Lippe, Rudolf (2005): *Weisheit - Wissen - Information*. Göttingen: V&R Unipress.
- Goldberg, David (1990): "Real-coded genetic algorithms, virtual alphabets and blocking". Urbana, Ill: Illinois Genetic Algorithms Laboratory.
- Goodman, Nelson (1978): *Ways of Worldmaking*. Indianapolis: Hackett Pub. Co.
- Görz, Günther; Rollinger, C.-R.; Schneeberger, J. (Hg.) (2003): *Handbuch der künstlichen Intelligenz*. 4. korr. Aufl. München [u.a.]: Oldenbourg.
- Gottschalk, Ralf (1999): „Abduktionen als Grund theoretischer und praktischer Dissense“. Zu finden auf uni-stuttgart.de
- Gottschalk-Mazouz, Nils (2012): "Toy Modeling. Warum gibt es (immer noch) sehr einfache Modelle in den empirischen Wissenschaften?" In: Peter Fischer, Andreas Luckner und

-
- Ulrike Ramming (Hg.): *Die Reflexion des Möglichen. Zur Dialektik von Handeln, Erkennen und Werten*: Lit. Verlag, S. 17–30.
- Grieser, G.; Fürnkranz, J. (2006): „Maschinelles Lernen. Symbolische Ansätze“. Darmstadt, 2006.
- Grimm, Jacob (1971): *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*. Band 16. 32 Bände. Leipzig: S. Hirzel.
- Grossekathöfer, Maik (15.03.2010): “Magnuson Carlsen on his chess career”. Interview mit Magnus Carlsen.
- Grossmann, Lev (2010): “How computers know what we want - before we do”. Zu finden auf time.com
- Harrach, Sebastian (2013): „Transklassischer Zugang zu Nichtwissen mittels maschinellern Lernen“. In: tbd (Hg.): *Proceedings to 23rd World Congress of Philosophy. Philosophy as Inquiry and Way of Life*. WCP2013. Athen: TBD.
- Hasler, Felix (2012): *Neuromythologie. Eine Streitschrift gegen die Deutungsmacht der Hirnforschung*. Bielefeld: transcript.
- Hassenzahl, Marc (2004): „Interaktive Produkte wahrnehmen, erleben, bewerten und gestalten“. In: F. Thissen und P. F. Stephan (Hg.): *Knowledge Media Design. Grundlagen und Perspektiven einer neuen Gestaltungsdisziplin*.
- Hayes, Jasmeet Pannu; Morey, Rajendra A.; Tupler, Larry A. (2012): “A case of frontal neuropsychological and neuroimaging signs following multiple primary-blast exposure”. In: *Neurocase: The Neural Basis of Cognition* (Volume 18, Issue 3), S. 258–269.
- Heidegger, Martin (1927): *Sein und Zeit*. I. Abteilung: Veröffentlichte Schriften 1914 - 1970. In: Martin Heidegger: Gesamtausgabe, Bd. 2. Hg. v. Friedrich-Wilhelm von Herrmann. Frankfurt am Main: Vittorio Klostermann.
- Heidegger, Martin; Herrmann, Friedrich-Wilhelm von (1989): *Gesamtausgabe. Die Grundprobleme der Phänomenologie*. 2. Auflage. Frankfurt am Main: Vittorio Klostermann (II. Abteilung: Volesungen 1923-1944, 24).
- Hippe, Seline (2012): „Software – Medium und kulturelle Technik“. Doktorandenkolloquium Lehrstuhl Prof. Hubig. Technische Universität Darmstadt. Darmstadt, 27.01.2012.
- Hoffman, Ralph E.; Grasmann, Uli; Gueorguieva, Ralitzia; Quinlan, Donald; Lane, Douglas; Miikkulainen, Risto (2011): „Using computational patients to evaluate illness mechanisms in schizophrenia“. In: *Biological psychiatry* 69 (10), S. 997–1005.
- Hubig, Christoph (2006): *Die Kunst des Möglichen I. Technikphilosophie als Reflexion der Medialität. Grundlinien einer dialektischen Philosophie der Technik*. Bielefeld: transcript.
- Hubig, Christoph (2007): *Die Kunst des Möglichen II. Ethik der Technik als provisorische Moral. Grundlinien einer dialektischen Philosophie der Technik*. Bielefeld: transcript.
- Hubig, Christoph (2008): „Der technisch aufgerüstete Mensch. Auswirkungen auf unser Menschenbild“. In: Alexander Rossnagel, Tom Sommerlatte und Udo Winand (Hg.): *Digitale Visionen. Zur Gestaltung allgegenwärtiger Informationstechnologien*. Berlin: Springer, S. 165–175.
- Hubig, Christoph (2008 II): „Technik als Medium und "Technik" als Reflexionsbegriff“. Philosophisches Kolloquium. Universität Stuttgart, 09.07.2008 II.
- Hubig, Christoph (2011): „Mythos und Religion als Institutionen. (Gehlen/Durckheim)“. Mythos und Rationalität. TU Darmstadt. Darmstadt, 03.05.2011.

-
- Hubig, Christoph (2012): „Technik - Natur - Kultur (und ihre Virtualisierung). Technik als Kultur“. TU Darmstadt. Darmstadt, 01.02.2012.
- Isermann, R.; Bender, E.; Bruder, R.; Darms, M.; Schorn, M.; Stählin, U.; Winner, H. (2009): „Antikollisionssystem PRORETA. Integrierte Lösung für ein unfallvermeidendes Fahrzeug“. In: Hermann Winner, Stephan Hakuli und Gabriele Wolf (Hg.): *Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort*. 1. Aufl. Wiesbaden: Vieweg, F, S. S. 632–646.
- Jones, Steve (2002): *Wie der Wal zur Flosse kam. Ein neuer Blick auf den Ursprung der Arten*. Ungekürzte Ausg. München: Dt. Taschenbuch-Verl.
- Kaiser, Lukasz (2012): “Learning Games from Videos Guided by Descriptive Complexity”. LIAFA, CNRS & Université Paris Diderot. Paris.
- Kakas, A.; Kowalski, R.A.; Toni, F. (1993): “Abductive Logic Programming”. In: *Journal of Logic and Computation* (2), S. 719–770.
- Kakas, A.; Kowalski, R.A.; Toni, F. (1998): “The role of Abduction in Logic Programming”. In: D. et al. Gabbay (Hg.): *Handbook in Artificial Intelligence and Logic Programming*, Bd. 5 (5), S. 235–324.
- Kaminski, Andreas (2010): *Technik als Erwartung. Grundzüge einer allgemeinen Technikphilosophie*. Bielefeld: transcript.
- Kaminski, Andreas (2012): „Lernende Maschinen: naturalisiert, transklassisch, nichttrivial? Ein Analysemodell ihrer informellen Wirkungsweise“. In: TBD.
- Kaminski, Andreas; Harrach, Sebastian (2010): „Do abductive machines exist? Proposal for a multi-level concept of abduction“. In: Klaus Mainzer (Hg.): *ecap10 - VIIIth European Conference on Computing and Philosophy*. München: Verlag Dr. Hut, S. 482–487.
- Kaminski, Andreas; Winter, Stefan (2011): „Paradigmenwechsel ohne Revolution: Ubiquitous Computing als Steigerungstechnologie. Zu einigen Kategorien der Technikgeschichte“. In: *Technikfolgenabschätzung – Theorie und Praxis* (3 (20)), S. 71–79.
- Kaplan, Abraham (1964): *The Conduct of Inquiry. Methodology for Behavioral Science*. San Francisco: Chandler Publishing Co.
- Karafyllis, Nicole C. (2003): *Biofakte. Versuch über den Menschen zwischen Artefakt und Lebewesen*. Paderborn: Mentis.
- Kasparov, Garry (2003): “Man vs. Machine - A new era in computer chess. Opinion by Garry Kasparov”. In: *The Wall Street Journal*.
- Kasparov, Garry (2010): *How Life Imitates Chess. Making the Right Moves, from the Board to the Boardroom*. Bloomsbury USA.
- Kerhet, A.; Small, C.; Quon, H.; Riauka, T.; Schrader, L.; Greiner, R. et al. (2009): “Application of machine learning methodology for pet-based definition of lung cancer”. In: *Current Oncology* (17), S. 41–47.
- Klöppel, Thomas; Popp, Alexander; Küttler, Ulrich; Wall, Wolfgang (2011): “Fluid–structure interaction for non-conforming interfaces based on a dual mortar formulation”. ISSN 0045-7825. In: *Computer Methods in Applied Mechanics and Engineering* (Volume 200, Issues 45–46), S. 3111–3126.
- Kogge, Werner (2008): „Technologie des 21. Jahrhunderts. Perspektiven der Technikphilosophie“. Literaturessay. In: *Deutsche Zeitschrift für Philosophie* 2008 (Vol. 56, No. 6), S. 935–956.

-
- Kononenko, Igor (2001): "Machine Learning for Medical Diagnosis. History, State of the Art and Perspective". In: *Artificial Intelligence in Medicine* (Volume 23, Issue 1), S. 89–109.
- Koza, John; Bennett, Forrest; Andre, David; Keane, Martin (1996): "Four problems for which a computer program evolved by genetic programming is competitive with human performance". In: *Proceedings of 1996 IEEE International Conference on Evolutionary Computation (ICEC '96)*. Piscataway, NJ: IEEE Service Center, S. 1–10.
- Kriesel, David (2007): „Ein kleiner Überblick über Neuronale Netze“. Zu finden auf dkriesel.com
- LeCun, Yann (zuletzt geprüft am 26.02.2013): „LeNet-5“. Zu finden auf yann.lecun.com.
- Mattern, Friedemann (2003): *Total vernetzt. Szenarien einer informatisierten Welt*; 7. Berliner Kolloquium der Gottlieb Daimler- und Karl Benz-Stiftung. Berlin [u.a.]: Springer.
- Mattern, Friedemann (Hg.) (2007): *Die Informatisierung des Alltags. Leben in smarten Umgebungen*. 1. Aufl. Berlin: Springer.
- Milner, Robin; Stepney, Susan (2003): "Nanotechnology: Computer Science opportunities and challenges". Submission by the UK Computing Research Committee to the Nanotechnology Working Group of the Royal Society and the Royal Academy of Engineering.
- Mitchell, Tom M. (1997): *Machine Learning*. New York: McGraw-Hill.
- Mooney, Raymond J. (2000): "Integrating Abduction and Induction in Machine Learning". In: Peter A. Flach und Antonis C. Kakas (Hg.): *Abduction and induction. Essays on their relation and integration*. Dordrecht, Boston: Kluwer Academic Publishers, S. 181–191.
- Nordmann, Alfred (2008): "Technology Naturalized. A Challenge to Design for the Human Scale". In: Pieter Vermaas, Perter Kroes, Andrew Light und Steven Moore (Hg.): *Philosophy and Design. From Engineering to Architecture*. Dordrecht: Springer, S. 173–184.
- Nordmann, Alfred; Bensaude-Vincent, Bernadette; Loeve Sacha; Schwarz, Astrid (2011): „Science vs. Technoscience. A Primer“. Version 2.0, Dezember 2011.
- Paavola, Sami (2006): On the Origin of Ideas. "An Abductivist Approach to Discovery". Philosophical Studies from the University 15. University of Helsinki, Faculty of Arts, Department of Philosophy.
- Peirce, Charles S. (1878): "Deduction, Induction, and Hypothesis". In: *The Popular Science Monthly* (13), S. 470–482.
- Peirce, Charles S.; Houser, Nathan (op. 2010): "Writings of Charles S. Peirce. A chronological edition". The architecture of theories. In: *Writings of Charles S. Peirce*.
- Peirce, Charles S.; Walther, Elisabeth (1991): *Vorlesungen über Pragmatismus*. [2. Aufl.]. Hamburg: Meiner (Philosophische Bibliothek).
- Rádio Xadrez (31.10.2012): „Rádio Xadrez interviews GM Boris Gelfand“. Interview mit Boris Gelfand.
- Raimer, Stephan (2002): „Anbahnung und Optimierung konstruktionsmethodischer Lernmuster mit hypermedialen Lernsystemen in der technischen Bildung“. Kiel.
- Rennie, Jason (2001): "Improving multi-class text classification with Naive Bayes".
- Rheinberger, Hans-Jörg; Herrgott, Gerhard (2001): *Experimentalsysteme und epistemische Dinge. Eine Geschichte der Proteinsynthese im Reagenzglas*. Göttingen: Wallstein-Verlag.

-
- Richter, Philipp; Kaminski, Andreas (2013): „Pragmatistische Eigentätigkeit der Mittel in Sein und Zeit“. Darmstadt, 28.03.2013. Gespräch mit Sebastian Harrach.
- Roco, Mihail; Bainbridge, William (2003): „Converging Technologies for Improving Human Performance“.
- Rogoff, Kenneth (2010): “Grandmasters and Global Growth”. Hg. v. Project Syndicate.
- Rosenblatt, Frank (1962): *Principles of neurodynamics. Perceptrons and the theory of brain mechanisms*. Washington: Spartan Books.
- Russell, Stuart J.; Norvig, Peter; Canny, John F. (2007): „Künstliche Intelligenz. Ein moderner Ansatz“. München [a.]: Pearson Studium (Informatik).
- Salomon, Ralf (1995): “Reevaluating Genetic Algorithm Performance under coordinate Rotation of Benchmark Functions. A survey of some theoretical and practical aspects of genetic algorithms”. In: *BioSystems*, Elsevier (39), S. 263–278.
- Scherer, Michael (2012): “Inside the Secret World of the Data Crunchers Who Helped Obama Win”. Zu finden auf swampland.time.com.
- Schmidt, A. (2007): „Eingebettete Interaktion – Symbiose von Mensch und Information“. In: Friedemann Mattern (Hg.): *Die Informatisierung des Alltags. Leben in smarten Umgebungen*. 1. Aufl. Berlin: Springer, S. 77–101.
- Scott, Jay (2001): “The neural net backgammon programs”. Zu finden auf satirist.org
- Sesink, Werner (2011): „Kompetenz in Technik“. In: Andreas Kaminski, Max Mühlhäuser, Werner Sesink und Jürgen Steimle (Hg.): *Interdisciplinary Approaches to Technology Enhanced Learning. Interdisziplinäre Zugänge zu technologie-gestütztem Lernen*. Münster: Waxmann.
- SETI Institute (2013): „FAQ. (Frequently Asked Questions)“. SETI Institute.
- Shah, Vatsal (2007): “Machine Learning Techniques for Stock Prediction”. Courant Institute of Mathematical Science, New York University.
- Shannon, Claude Elwood (1949): “Programming a computer for playing chess”. Murray Hill, N.J.
- Shiffman, Daniel (2012): *The nature of code*.
- Smith, Kenneth (2009): “Morphing Wing Fighter Aircraft Synthesis/Design Optimization”. Virginia Polytechnic Institute and State University.
- Spix, Jörg (1998): „Maschinelles Lernen. Analytisches Lernen“. Fachbereich 10 der Universität Oldenburg. Oldenburg, 31.03.1998.
- Stallkamp, Johannes; Schlipsing, Marc; Salmen, Jan; Igel, Christian (2011): “The German Traffic Sign Recognition Benchmark. A multi-class classification competition”.
- Teuchert-Noodt, Gertraud (2011): „Wie das Lernen funktioniert. Neue Ergebnisse aus der Hirnforschung. Didaktik der Informationstechnik“. Technische Universität Darmstadt. Darmstadt, 28.11.2011.
- Tönnsen, Kai-Christian (2007): „Potentialerweiterung webbasierter und hypermedialer Lernsysteme durch Integration technischer Experimente und Realobjekte“.
- Tsinteris, Kimon; Wilson, David (Zuletzt geprüft am 15.03.2013): “TD-learning, neural networks and backgammon”. Zu finden auf cs.cornell.edu.
- Weber, Tassilo (2010): *Der Weltbegriff in Heideggers Fundamentalontologie und Nelson Goodmans Symboltheorie*: GRIN Verlag.

-
- Wei-Ding, Tsai (2011): „Die ontologische Wende der Hermeneutik. Heidegger und Gadamer“. Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophie an der Ludwig-Maximilians-Universität München.
- Weimer, Markus (2010): „Machine Learning und Erfahrungen aus der Industrie“. Robert-Piloty-Gebäude, A126, Darmstadt, 23.09.2010.
- Weimer, Markus; Karatzoglou, Alexandros; Smola, Alex (2008): “Adaptive Collaborative Filtering”. In: Pearl Pu, Derek Bridge, Bamshad Mobasher und Francesco Ricci (Hg.): *Proceedings of the 2008 ACM conference on Recommender systems. RecSys 2008*. Lausanne, Switzerland, October 23-25. New York: Association for Computing Machinery.
- Wiegerling, Klaus (2011): *Philosophie intelligenter Welten*. Paderborn: Fink.
- Wiegerling, Klaus (2012): „Der neue Mensch zwischen Cyborg und Biofakt“, 11.07.2012.
- Wiegerling, Klaus; Heesen, Jessica; Siemoneit, Oliver; Hubig, Christoph (2008): „Ubiquitärer Computer - Singulärer Mensch“. In: Dieter Klumpp, Herbert Kubicek, Alexander Roßnagel und Wolfgang Schulz (Hg.): *Informationelles Vertrauen für die Informationsgesellschaft*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, S. 71–84.
- Wikibooks Contributors (zuletzt geprüft am 07.11.2011): “Martin Heidegger/ Sein und Zeit”. In: Wikibooks. Florida: Wikimedia Foundation, Inc.
- Wikipedia Contributors (zuletzt geprüft am 07.11.2011): “100-Schritt-Regel”. In: Wikipedia. (deutschsprachig). Florida: Wikimedia Foundation, Inc.
- Wikipedia Contributors (zuletzt geprüft am 05.03.2013): “Gregory Bateson”. In: Wikipedia. (deutschsprachig). Florida: Wikimedia Foundation, Inc.
- Wikipedia Contributors (zuletzt geprüft am 07.01.2013): “Go. (Spiel)”. In: Wikipedia. Florida: Wikimedia Foundation, Inc.
- Wikipedia Contributors (zuletzt geprüft am 13.02.2012): “Induktive logische Programmierung”. In: Wikipedia (deutschsprachig). Florida: Wikimedia Foundation, Inc.
- Wikipedia Contributors (zuletzt geprüft am 07.11.2011): “Information”. In: Wikipedia (deutschsprachig). Florida: Wikimedia Foundation, Inc.
- Wikipedia Contributors (zuletzt geprüft am 07.11.2011): “Konnektionismus”. In: Wikipedia (deutschsprachig). Florida: Wikimedia Foundation, Inc.
- Wikipedia Contributors (zuletzt geprüft am 07.11.2011): “Neuronales Netz”. In: Wikipedia (deutschsprachig). Florida: Wikimedia Foundation, Inc.
- Wikipedia Contributors (zuletzt geprüft am 10.10.2012): „OLAP-Würfel“. In: Wikipedia (deutschsprachig). Florida: Wikimedia Foundation, Inc.
- Wirth, Uwe (1995): „Abduktion und ihre Anwendung“. In: *Zeitschrift für Semiotik* 1995 (17), S. 405–424.
- Wirth, Uwe (2003): „Die Phantasie des Neuen als Abduktion“. In: *Deutsche Vierteljahresschrift* 77 2003 (4), S. 591–618.
- Zheng, Yan-Tao; Zhao, Ming; Song, Yang; Adam, Hartwig; Buddenmeier, Ulrich; Bissacco, Alessandro et al. (2009): “Tour the World: building a web-scale landmark recognition engine”. Google Inc. U.S.A.